Categorization

Douglas L. Medin
Evan Heit

I. OVERVIEW

A. Current State of Affairs

The study of categorization is not unlike the cooperative algorithms associated with neural net models. That is to say, there is a synergy between the various areas associated with cognitive science that makes categorization research dynamic and exciting. Individual subdisciplines have matured to the degree that they are interacting with one another more extensively than perhaps ever before. Consider the question of the extent to which categorizing in any particular situation is based on the application of abstract rules or on the retrieval of specific memories. For example, is categorizing a novel item (e.g., "X is an instance of Disease Y") driven by checking the consistency with abstract rules or deciding that the item is sufficiently similar to a remembered instance of a category (e.g., Rips, 1989; E. E. Smith & Sloman, 1994)? It turns out that memory-based categorization in cognitive psychology is closely related to a development in artificial intelligence (AI) known as case-based reasoning. And the question of memories versus rules is highly relevant to issues in philosophy concerning concept stability. Finally, the long-held belief in linguistics that knowledge of a grammar is represented by linguistic rules has been challenged by connectionist researchers who hold that knowledge is far more specific than rules. In short, there are strong motivations for mutual interaction across traditionally distinct research areas.

Cognitive Science

Of course, the various cognitive science subdisciplines approach a given issue from different, complementary perspectives. Our review of current work in categorization reflects these diverse methods of inquiry. It also reflects the fact that we are cognitive psychologists; cognitive scientists from other areas would no doubt have different patterns of emphasis and organization.

B. Brief History

One of the most central questions in categorization concerns the structure of concepts. By *concept* we mean a mental representation of a category serving multiple functions, one of which is to allow for the determination of whether or not something belongs to a class. Following E. E. Smith and Medin (1981), one can distinguish three positions on conceptual structure. The *classical view* holds that all instances of a category share common properties that are necessary and sufficient conditions for defining the concept. The *probabilistic view* denies that there are defining properties and instead claims that concepts are organized in terms of properties that are only characteristic of category instances. Membership in a category can thus be graded rather than all-or-none, where the better members have more characteristic properties than the poorer ones. Perhaps the most exciting development in the psychology of concepts in the 1970s was the shift from the classical to the probabilistic view, importantly motivated by Eleanor Rosch's studies of natural object categories (see Mervis & Rosch, 1981; Medin & Smith, 1984, for reviews, and Margolis, 1994, for a recent critique; E. Rosch, 1978; e.g., E. Rosch & Mervis, 1975; E. E. Smith, Shoben, & Rips, 1974). The third view of conceptual structure, the *exemplar view*, agrees with the claim that concepts need not contain defining properties, but further claims that categories may be represented by their individual exemplars, and that classification is determined by whether the instance is sufficiently similar to one or more of the category's known exemplars (e.g., Brooks, 1978; Medin & Schaffer, 1978).

It is obvious that views of conceptual structure constrain models of category processing. For example, hypothesis testing or rule-based models are more compatible with classically defined than with probabilistic categories. Hypotheses would correspond to conjectures about the defining properties, and if there are no defining properties, any conjunctive rule would need to have a procedure for dealing with exceptions. A learning procedure that abstracts the central tendency of category examples, such as prototype formation (e.g., Posner & Keele, 1968), is especially compatible with the probabilistic view. In fact the probabilistic view is sometimes referred to as the prototype view. At a more general level, all three views of conceptual structure are consistent with similarity-based models of category learning. For purposes of illustration we could assume some generic similarity model where similarity is some weighted function of shared and distinctive properties. Classical view category structures would correspond to the special case of the similarity model, where the defining features receive all the weight. Probabilistic view category structures could be learned as a system that also weights important or characteristic features regardless of whether or not they are defining. Finally, the exemplar view explicitly assumes that similarity to examples *is* the processing system determining classification. We belabor this point about similarity because similarity-based models of category learning have recently come under criticism (e.g., Murphy & Medin, 1985; Schank, Collins, & Hunter, 1986).

One challenge for similarity-based models of categorization is adequately constraining the notion of similarity. According to Tversky's (1977) contrast model, for example, similarity is a weighted function of matching and mismatching features. Therefore, the similarity between two entities depends crucially on which features enter into comparison for matches and mismatches and on the weights assigned to the features being compared. Potentially, then, similarity is an empty notion devoid of explanatory power (Goodman, 1972). To make similarity meaningful, there must be constraints on the features and their weighting. In that event, however, it is the constraints that are performing the explanatory work, not the abstract notion of similarity (e.g., Medin, 1989). Of course, one might argue that the human perceptual system has evolved to select the right kinds of similarity, namely, those that give useful categorization schemes (Murphy & Medin, 1985). The success of similarity-based learning models suggests that this is at least part of the story.

Another source of constraints might be theories and other forms of knowledge that pick out and weight relevant features. In AI this approach to categorization is referred to as explanation-based learning (e.g., Dejong, 1988). The idea is that prior knowledge in the form of a domain theory can be used to explain *why* some example is a member of a category and then can be used to generalize the concept appropriately. For example, if the explanation for why some example is a *cup* does not include the color of the cup as part of the explanation, then the generalization will not contain color as a relevant feature (Mitchell, Keller, & Kedar-Cabilli, 1986). Therefore, if the example were a red cup, the system would show no tendency to act as if other red things might be cups.

Later on we shall have much more to say about both similarity-based learning (SBL) and explanation-based learning (EBL). For now we simply note that a great deal of current research and theory in categorization is directed both at contrasts between these two approaches and, more significantly, at ways of integrating or combining them.

C. Ecological Validity and Artificial versus Natural Categories

Anyone who starts reading psychological research on categorization will quickly notice that studies using natural object categories and those using artificially constructed categories are both prominently represented. By natural object categories we mean categories with lexical entries (e.g., *bird*) whose instances correspond to entities in the world (e.g., *robin, turkey, pigeon*). Artificially constructed categories typically involve novel stimuli where the constituent features or properties of examples are familiar, but where the experimenter specifically manipulates the properties of examples and the assignment of examples to categories to create some particular category structure of interest. Neither the examples nor the categories need necessarily correspond to real-world entities.

There is both interplay and tension between work using natural categories and that using artificial categories (e.g., see Murphy, 1993, for a discussion). On the one hand, it seems straightforward that the closer an experimental situation is to real-world contexts, the more readily one may generalize to those contexts. Results from artificial categories may be apropos of nothing. A counter-argument is that real-world contexts are characterized by numerous correlated (and, therefore, confounded) variables and that artificial categories are needed to run properly controlled experiments capable of isolating variables relevant to categorization. We believe that both positions have validity; researchers may profitably and explicitly violate ecological validity for certain purposes, but they cannot ignore it. Let's take a quick look at two examples of an effective interplay between the artificial and the natural (see Medin & Thou, 1992, for further examples and discussion).

One general rationale for using artificial categories is to identify some variable or structural property of natural categories, incorporate that property into artificially constructed categories, and then conduct experiments to evaluate the role and importance of that property in categorization. This strategy allows one to control for a variety of extraneous variables that might affect performance with natural categories. Consider, for example, Rosch's pioneering research on goodness of example or typicality effects (e.g., E. Rosch, 1973; E. Rosch & Mervis, 1975). Rosch proposed that natural categories have a *family resemblance* structure giving rise to typicality effects. Under the family resemblance principle, category members considered the most typical are those with the most properties in common with other category members and the fewest attributes in common with members of contrasting categories. E. Rosch and Mervis (1975) employed a variety of measures that converged to suggest that natural categories have a family resemblance structure. They then created artificial categories according to a family resemblance principle, ran learning and category verification studies, and observed the same pattern of goodness-of-example effects that had been observed earlier. This replication reinforces the claim that family resemblance structure rather than some other factor, such as familiarity, is responsible for typicality results.

The interplay between the natural and the artificial can also flow in the other direction. For example, Lee Brooks and his colleagues (e.g., Allen & Brooks, 1991; Brooks, 1987; Reagher & Brooks, 1993) have used artificial stimuli and categories to study the influence of specific item similarity on categorization. One striking result is that even when participants are explicitly asked to employ a straightforward rule to categorize, their responses are influenced by irrelevant (from the perspective of the rule) similarities. That is, people are slower and less accurate at classifying an example as a noninstance of the rule if it shares rule-irrelevant similarities with specific examples that do instantiate the rule. These specific item similarity influences are interesting in their own right, but Brooks and his associates have also used natural categories to pursue the implications and generality of their results (e.g., Brooks, Norman, & Allen, 1991). In the domain of dermatology, Brooks et al. (1991) found that the diagnosis of skin disorder was facilitated by similar cases previously seen in the same context. In short, episodic influences on categorization are quite robust (see also Weber, Bockenholt, Hilton, & Wallace, 1993).

We believe that the above two instances are examples worth imitating. Without some concern with real-world contexts, there is always the risk that critical variables are being ignored (critics of similarity-based models of categorization might argue that the influence of prior knowledge on categorization is one such variable). But progress often requires controlled observations, and artificially created categories are an important tool for doing so.[1]

D. Summary

Although we have alluded to a parallel, interactive pattern of categorization research activity, we necessarily are constrained to describe it in a serial manner. Our review will be organized around methods of inquiry, and certain themes and issues will recur in virtually every section. We begin with evolutionary considerations and then turn to philosophical perspectives. Next, we review developmental and cross-cultural research, followed by a focus on computational models, and finally by observations from neuroscience. Our survey will necessarily not be comprehensive but we hope to at least convey the flavor and some of the strong points of each of these methodologies. In the last section of this chapter, our goal is to provide a summary and integration by discussing challenges and opportunities in categorization research.

II. METHODS OF INQUIRY

A. Evolutionary

Although human cognition presumably is adaptive, until recently cognitive psychologists have placed little emphasis on the purpose or function of cognitive activities. One reason for this neglect is that it is far from clear how one might go about testing or providing independent evidence bearing on hypotheses about function. Cognitive scientists such as Marr (1982), however, have demonstrated the value of multiple levels of analysis, including the broad question of what an intelligent system is trying to compute. In the domain of categorization one clear benefit of an

[1]Of course the distinction between artificial and natural categories can to some extent be blurred. Researchers may create artificial categories from natural stimuli (e. g., photographs of faces) or employ stimulus materials where prior knowledge comes into play (e. g., Wattenmaker, Dewey, Murphy, & Medin, 1986). Nonetheless, we think it is important to keep the potential trade-offs between artificiality and generalizability in mind in evaluating research and theory in categorization.

evolutionary perspective is the realization that concepts serve multiple functions and a focus on a single function comes at the risk of developing theories that are too narrow to do the work they ultimately will be asked to do (see Matheus, Rendell, Medin, & Goldstone, 1989, for applications in both AI and cognitive psychology). We begin with a brief summary of conceptual functions and some research directly linked to evolutionary analysis. Of course the reader will want to keep these functions in mind as they are also relevant to subsequent sections of this review.

1. Conceptual Functions

We distinguish eight distinct functions of concepts: categorization, understanding, learning, inference, explanation, conceptual combination, planning, and communication. By the *categorization* function of concepts we refer to the fact that mental representations are used to determine the category membership of entities. Indeed, psychologists are sometimes accused (perhaps correctly) of assuming that the *only* function of concepts is to classify, and that conceptual representations include little more than procedures for identifying category membership (e.g., Mandler, 1993). But categorization must be more than an end in itself.

Categorization is also a procedure for relating new to old. Even when objects or events are novel, the cognitive system is capable of bringing relevant knowledge to bear in the service of *understanding*. For example, as a result of categorizing some object as a *telephone*, people understand its relevant parts and how they might interact with it.

A critical conceptual function is to *support learning*. New entities are understood in terms of old, but the new also feeds back to modify or update the knowledge used in categorization. This broad function is itself associated with a variety of critical questions concerning adaptation and adaptiveness. For example, updating must balance the need to be relevant to contemporary contexts with the danger of discarding accumulated wisdom. Furthermore, given that instances partake of multiple category memberships (e.g., some white fury thing can be categorized as a poodle, a dog, a mammal, a pet, a domestic animal, and so on) should each experience modify all possible categories or just some relevant subset?

A fourth important function of concepts is *inference*. Having categorized some entity we can make predictions concerning its behavior. In the domain of medicine, diagnostic categories allow the physician to predict what sort of treatment might prove effective. As we shall see, there has been a recent upsurge of research and theory on category-based inferencing.

Concepts are critically involved in *explanation* and *reasoning*. Having categorized a young man who is cleaning a sidewalk with a toothbrush as a *fraternity pledge* one can provide a reason for his strange behavior. Furthermore, sometimes concepts may be used to persuade; we are all familiar, for example, with the use of labels in political campaigns.

By combining concepts we can use a limited number of concepts to create an unlimited number of new concepts. Just how people modify constituent concepts in order to comprehend combined concepts remains a significant challenge. For example, we understand *car repair* as repair of a car but *expert repair* as repair by an expert (Murphy, 1988).

Additionally, we use categories to instantiate goals in planning. For example, one might think about what sort of food one would like for dinner or what things one ought to bring along on a camping trip (e.g., Barsalou, 1983).

Finally, we use concepts for communication. The interpersonal aspect of concepts places constraints on virtually every other conceptual function. We will not cover this function in detail in this review, but we wish to make the distinction clear between communication and other functions. For example, a speaker might first categorize some object as a *Toyota Tercel,* then refer in conversation to this object as either a *vehicle,* a *car,* a *foreign car,* or a *Toyota,* depending on the speaker's communicative goals (Grice, 1957). (See Malt, 1990, 1994, for further discussion of the distinction between categorization and naming.)

These above functions may place competing demands on conceptual organization. For example, one can maximize prediction from category memberships by developing very narrow categories. The cost of this precision is that new examples may frequently fall into no preexisting category, undermining the understanding function of concepts. We turn now to some specific applications and ideas inspired by the evolutionary framework.

2. Purpose of Categorization

a. Anderson's Rational Model

One clear example of an evolutionary or ecological approach to categorization is Anderson's rational model (Anderson, 1990, 1991). Anderson argues that the human mind is a rational, (close to) optimal system, and that one can construct models that are excellent approximations to human performance by analyzing what is optimal. He assumes that what is being optimized in categorization is predictability of features or properties (i.e., what we have referred to as the inference function of concepts). The next step in his analysis is to describe the structure of information in the environment. This analysis leads to three main conclusions or hypotheses: (a) features are probabilistically associated with categories; (b) categories are a (nearly) disjoint partitioning of objects in the world; and (c) features or properties within a category are (approximately) independently distributed.

The third step in Anderson's rational analysis is to derive the optimal function, taking into account the costs associated with different functions. The ideal algorithm would keep track of all possible partitions to select the partition with maximum predictability. However, in the case of categorization, the optimal procedure cannot be run on a computer because there are too many possible partitionings (there's a combinatorial explosion for all but the smallest number of objects). As an alternative, Anderson developed an iterative algorithm in which members are considered incrementally and classified into the category that maximizes the predictability of the resulting partitioning (see Anderson, 1990, for details).

The prediction rule of the rational model works as follows: When some novel object is presented and the system has to predict whether some feature *k* is present, the algorithm calculates the probability that the object is in each of its categories multiplied by the probability of each category having feature *k*. The sum of this cumulative function is the predicted probability that the novel object has feature *k*.

Anderson's rational model is quite successful at predicting the results of a number of category learning experiments using artificial categories. To apply the model to the classification of examples, one treats the category label simply as another feature that one is trying to predict.

Is Anderson's rational model superior to alternative categorization models? For a variety of reasons the answer is not obvious. Nosofsky (1991) provided a formal proof that the rational model is a generalization of an exemplar model of categorization, the Medin and Schaffer (1978) context theory. Nosofsky examined 11 sets of transfer data from categorization experiments and found that the general form of the rational model failed to provide an account of transfer superior to the special case corresponding to the context model. Heit (1992) extended the context theory to prediction and inference tasks (rather than just categorization) and to reasoning using chains of examples. This extended context model successfully predicted the results of several experiments in which subjects made predictions about transfer stimuli. Heit also applied the rational model to these data on inferences, and it provided no improvement over the multiple-step context model.

As Murphy (1993) notes, the rational model does not make much use of its categories in prediction in that it sums categorization and feature probabilities over all categories. Ahn and Medin (1992) directly evaluated the rational model's predictions concerning category construction. They presented people with sets of examples and asked them to create categories, in some cases according to certain constraints (e.g., two categories of equal size). The examples had a feature structure that could be used to create family resemblance categories. Ahn and Medin observed family resemblance sorting under some conditions, but the rational model was unable to predict when family resemblance sorting would or would not occur. The results were consistent with a two-stage model, which assumes that people impose more structure than the examples support in the first stage, and that the second stage adjusts for this difference between perceived and preferred structure (see also Medin, Wattenmaker, & Michalski, 1987, for analogous results in a rule induction task).

Thau (1992) also has evidence that people actively organize categories. He employed an incremental clustering task where instances are presented one at a time; a key experimental manipulation was the order in which the examples were presented. His goal was to create a situation where two distinct orders would yield the identical category structure after *n* examples. The question was whether the *n* + 1 example would be categorized the same way for the two orders. Many incremental clustering models, including the rational model, predict no effect of order.

Thau observed clear order effects that were consistent with the idea that selective attention or weighting is given to dimensions used to organize early examples. Again this is consistent with active organization rather than an unbiased weighting of probabilistic structures.

b. Critique

Anderson's rational model illustrates both the potential strengths and shortcomings of an evolutionary or ecological analysis. Each step in the analysis offers challenges and potential problems. For example, although prediction is an important function of concepts, we would argue that it is far from being the only one. Murphy (1993) provides a detailed criticism of each of Anderson's assumptions concerning the structure of information in the environment. Murphy points out that categories (a) often have multiple overlapping membership functions (e.g., *dog, pet*) rather than being disjoint; (b) that questions about feature structure may be meaningless unless one can provide constraints on features; and (c) that features within a category may be correlated rather than independent (e.g., Malt & Smith, 1984). Finally, Murphy argues that because optimal functions need to consider computational constraints, one needs a fairly concrete process model, as different models for the same task can require very different constraints.

The upshot of these criticisms is that an ecological analysis is difficult—not that it should not or cannot be done. Indeed, Murphy (1993) argues in favor of an ecological analysis that focuses on real-world concept learning and use. He suggests that such analysis will include attention to the learner as well as to the environment by itself. For example, conceptual domains may be organized by theories and may change with expertise, according to Murphy.

c. Shepard's Universal Law of Generalization

As another illustration of the evolutionary approach to concepts, we will briefly describe work by Shepard (1987). Shepard argues that all organisms are faced with the task of *generalization;* because no stimulus ever recurs in exactly the same way, organisms constantly have to categorize novel stimuli and make other inferences about them. In Shepard's terminology, a learner has to decide what is the *consequential region* for an observed stimulus. In other words, after observing an object located at some point in multidimensional psychological space, the learner must assume that other objects within an enclosed region of psychological space near the first object will have the same consequences as that original object. (For example, if a bird eats a caterpillar then gets sick, the bird may generalize that other caterpillars of similar color and markings are also poisonous.) From these basic assumptions about categorization and prediction, Shepard derived a mathematical account of the functional form of generalization in humans and other species. In particular, for a variety of possible shapes of consequential regions, and regardless of the size of the consequential region, the probability of generalizing from an old observation to a new observation is a negative exponential function of the psychological distance between the two stimuli. Shepard (1987) reviewed many studies on generalization that indeed showed just this functional form. Importantly, some recent models of categorization make predictions about generalization similar to Shepard's evolutionary account (Gluck, 1991).

d. Summary

Clearly the rational analysis by Anderson has been valuable in serving as the basis for analysis of critical issues in categorization. The contrast between Anderson and Murphy illustrates the differences in perspective of similarity-based and explanation-based approaches to categorization. A key question is whether we have these concepts because of the way the world is structured or because of the way human minds are structured. Likewise, Shepard's account of categorization and prediction was derived by considering the basic tasks faced by any organism, and notably Shepard's account is similar to descriptive psychological models of categorization. We think the observation of William James (1890) is most apt: "mind and world have evolved together, and in consequence are something of a mutual fit" (p. 47). Within the broad spectrum of that something of a fit, however, is room for a great deal of exciting debate and interaction. Asking and addressing questions about conceptual functions has been and, in our opinion, will continue to be a key organizing principle for categorization research.

B. Philosophical

The contributions of philosophical perspectives to categorization research have been substantial and have taken a variety of forms. For present purposes we will focus on some observations concerning reference and meaning and their implications for conceptual structure.

1. Reference

Psychologists have tended to assume that concept representations mediate the link between words and the things to which they refer. For example, one's concept of *triangle* contains information that determines reference—the things that are or are not triangles. Putnam, however, has argued that words may refer more directly (see Rey, 1983). Borrowing an example from Komatsu (1992), a person who knows nothing about *sassafras* may ask, "What is sassafras?" thereby referring to sassafras, without having any conceptual representation of it. If a mental representation is not necessary as a mediator in this example, direct reference may also be possible in situations where a conceptual representation is available (e.g., when one is somewhat familiar with sassafras). Putnam claims that for some types of nouns, such as natural kind terms (e.g., *gold, water,* biological kinds), individual people's conceptual representations do *not* establish reference. Instead, the reference of natural kind terms is ultimately a matter of discovery. That is, it is a matter for science to determine the true nature of things like *gold* and *water.*

Putnam's view has attracted a great deal of attention and debate. One issue concerns just which kinds of concepts entail direct reference (e.g., Dupré, 1981, Losonsky, 1990; Schwartz, 1977). Malt (1990) examined the acceptability of linguistic hedges as a means of evaluating people's beliefs about categories. The hedges included phrases like "according to experts," "technically speaking," "by definition," and "loosely speaking." Her results are consistent with the idea that concepts may differ from one another not only in the properties represented in the concept, but also in the belief held about the completeness and validity of these properties as a description of their referents. For instance, the hedge "according to experts" was much more acceptable for natural kind examples than for artifacts, whereas the reverse held for the hedge, "loosely speaking." Malt's observations are consistent with the idea that theories and beliefs play a role in conceptual organization not captured by constituent properties per se.

In related work, cognitive psychologists have shown that judgments of category membership are critically affected by theories about the underlying natures of entities. For example, Rips (1989) gave people scenarios where an animal changed from having bird-like properties to having insect-like properties so that, perceptually, it became much more like an insect than a bird. The transformation was framed in terms of normal development or as a response to hazardous chemicals. People judged that animal from the first stage more likely to be truly a bird when the transformation was an accident than when it was part of normal development (see also Keil, 1989).

Researchers have also worried about the extent to which laypeople do defer to science. Although we are willing to categorize whales as mammals despite their superficial similarity to fish, we continue to employ our concept of *tree* even though taxonomists point out that trees do not comprise a kind. Malt (1994) has studied the set of fluids that people call *water.* Contrary to Putnam, she finds that although some amount of H_2O may be necessary for something to be called water, the percentage of H_2O in a liquid is a rather poor predictor of what people will call water.

If reference can change as our beliefs change, then what gives our concepts stability? (See Rey, 1983, E. E. Smith, Medin, & Rips, 1984, for a more detailed discussion of this issue.) Rips (in press) proposes a distinction between mental representations *of* a category from mental representations *about* a category. He suggests that the former need only consist of node or marker, functioning to provide both intra- and interindividual concept stability. The part of the representation that is about the category consists of theories and beliefs that themselves may change within an individual or be different across individuals.

2. Kinds of Categories

Another important aspect of Putnam's analysis is the idea that there are distinct *kinds* of categories. Putnam maintains that the reference of natural kind terms is based on underlying natures and is a matter of discovery. In contrast, the reference of nominal kind terms such as *bachelor* is established by convention (see also Donnellan, 1977). Artifact categories may be intermediate in their dependence on convention versus discovery. Although these distinctions may involve more of a continuum than a dichotomy, they do emphasize two important dimensions of conceptual structure that are central to theories of conceptual change. Consider the situation where some entity is categorized as belonging to some category and that the entity then manifests some surprising property or properties. At this point there is tension between modifying one's concept versus one's belief that the entity is, in fact, a member of that category.

3. Meaning

To the extent that reference is a matter for experts to determine and to the extent that meaning includes relations between concepts and referents, the study of concepts does not provide a full account of meaning (Fodor, 1983). To clarify this point we need to distinguish between metaphysics and epistemology. Metaphysics is concerned with issues about how the world is, whereas epistemology is concerned with how we know, believe, or infer that the world is. Reference for natural kinds is presumably a matter of metaphysics; psychological studies of categorization are confined to epistemological questions. For example, the classical view of concepts represents a claim about mental representations of categories, not the categories themselves. Things in nature may have an underlying nature that people remain ignorant of, or conversely, people may believe that certain concepts (e.g., of a species or biological kind) have a shared underlying essence even though science fails to support this view (e.g., Mayr, 1982).

4. Summary

Our treatment of the philosophical perspective on categorization research necessarily has been limited. We hope it is clear that questions about relationships among category representations, reference, and meaning are central to our understanding of categorization. The possibility that there are distinct kinds of categories is one important by-product of these analyses.

C. Developmental

Developmental research has had a truly major impact on the psychology of categorization. Questions about learning and the role of similarity versus theories in conceptual organization are sharply focused in developmental studies. The chapter 5 (this volume) by Carey and Markman describes much of this body of work, and we will only touch on a few high points.

1. Similarity and Development

A number of researchers have suggested that there are developmental changes in similarity processing that have straightforward linkages to categorization. One way to characterize the shift is to propose that young children process stimuli holistically and that older children are more analytic (e.g., Kemler, 1982, 1983; Kemler-Nelson, 1988; Shepp & Swartz, 1976; L. B. Smith & Kemler, 1977, 1978; but see also Ward, Vela, & Hass, 1990). Linda Smith (1989) offered a mathematical model of similarity processing where the central claim was that relative to younger children, older children are more likely to selectively weight dimensions and more likely to give greater weight to matching values on a dimension.

Gentner and Ratterman (1991) argue for a shift from processing focusing on simple attributes or properties (one-place predicates: e.g., "X is green") to attention to relational properties (multiplace predicates: e.g., "X is above Y" or "X causes Y"). Keil and Batterman (1984) report a developmental shift in categorization from a reliance on characteristic features to a greater emphasis on more central or defining characteristics. These two sets of observations converge if more central properties tend to be more relational in character.

These various distinctions all seem to point toward a story whereby younger children employ SBL that is later integrated with theory-based learning. Theories seem to involve analytical processes, selective weighting, and relational predicates or properties, each of which seems to appear later in development. But this is only part of the story, and many of the authors we have cited would disagree with it. Let's take a closer look.

2. Constraints, Theories, and Development

Keil (1989) refers to the above story about SBL being followed (developmentally) by theory-based reasoning as the "doctrine of original sim" and he rejects it. Instead, he claims that even very young children's category learning is constrained by domain-specific theoretical biases. He is not alone in this view. Carey (1985) argues that there is a development shift in children's biological reasoning from being organized in terms of a naive psychology to being based on a naive biology. For example, she has evidence that young children's inductive inferences are guided by similarity to humans, not an unbiased overall similarity. That is, they are more confident that a *bee* has some novel property if they are told that *people* have it than if they are told *bugs* have it.

There is evidence that babies are sensitive to the animate–inanimate distinction (e.g., Leslie, 1984) and that young children use animacy as an organizing principle even when it conflicts with overall similarity (R. Gelman, 1990; S. A. Gelman & Wellman, 1991). In short, it does not appear that there is a stage of development where similarity has the turf to itself.

Of course, researchers do not necessarily assume that children hold theories that have all the properties of scientific theories. So what do they mean? As one example, let's look at the notion of *psychological essentialism* (e.g., Atran, 1990; S. A. Gelman & Wellman, 1991; Keil, 1989; Medin & Ortony, 1989). The main idea is that (at least for the domain of biological things and events) people act as if things in the world have a true underlying nature that imparts category identity. Furthermore, this essence is thought to be the causal mechanism that generates visible properties. Therefore, surface features provide clues but are not infallible indicators of category membership. We refer to this view as *psychological* essentialism because it is concerned with people's *assumptions* about how the world is, not how the world truly is.

Developmental psychologists (especially Susan Gelman and Frank Keil) have provided evidence consistent with the view that children reason in an essentialist manner (S. A. Gelman, Coley, & Gottfried, 1994; S. A. Gelman & Wellman, 1991; Keil, 1989; Springer & Keil, 1991). Four types of evidence seem especially relevant: (a) appeal to invisible causal mechanisms to explain appearance and changes associated with growth; (b) the assumption of innate dispositions or inborn capacities to explain capacities that emerge later in life; (c) belief in the maintenance of identity despite changes in superficial appearance; and (d) the assumption that members of a category share a large number of other properties that may be hidden or unknown. S. A. Gelman et al. (1994) review extensive supportive evidence for each of these assumptions (see also Shipley, 1993, for a cogent discussion of induction potential).

There is clear evidence that category labels are treated by children as especially significant for reasoning. For example, S. A. Gelman and Markman (1986) pitted category membership against perceptual similarity in an inductive reasoning task. Young children were first shown pictures of two animals and taught that different novel properties were true of them. Then they were asked which property was true of a pictured new example. The new example was perceptually similar to one of the first pictures, but it shared category membership with the other (which was not similar to the new example). Children judged that the new example would have the property of the animal that was of the same category but perceptually different. Even for young children, similarity acts as a general guideline that can be overridden by other forms of knowledge.

This emphasis on the role of theories in category development is not a universal assumption. For a recent critique see S. S. Jones and Smith (1993) and the associated commentaries (Barsalou, 1993; S. A. Gelman & Medin, 1993; Mervis, Johnson, & Scott, 1993).

3. Category Labels and Learning

Our discussion so far has treated category learning as straightforward and ignored some difficult problems associated with learning the referent or extension of concepts. As Quine (1960) noted in his discussion of translation, there is an inherent ambiguity in reference. When a parent points to a rabbit and says "rabbit" the situation is much more complex than might first appear. The word could refer to "small" or "white," or "furry" or "hopping," or a proper name or "pet" or "animal," or any of an unlimited number of other things. The learning situation is seriously underconstrained. So how do children learn what's what?

Ellen Markman (1989, 1990) has suggested three constraints that children place on reference and provided empirical support for each of them. One bias is that children tend to assume that a novel term applies to the entire object rather than to its parts or properties. Another is that children act as if they expect objects to have only one label. Note that this assumption is objectively incorrect, but may nonetheless be helpful in initial stages of language learning. For example, suppose a child sees two objects and hears a novel label. If the child knows the label for one of the objects he or she might assume that the novel label applies to the other object (Markman, 1992; Markman & Wachtel, 1988). A third bias is to assume that labels refer to objects of like kind rather than to objects that are contextually or thematically related. In related work, Landau, Smith, and Jones (1988) have found that children's assumptions about "like kind" shift from overall similarity to shape similarity as one moves from a nonlinguistic context to a linguistic context.

4. Kindhood and the Basic Level

Earlier we mentioned that a given entity may have numerous category memberships. One may ask whether any of these is privileged with respect to notions of like kind or kindhood. Rosch's innovative studies of natural object concepts included an analysis of the vertical component of categories. For example, an object being driven down the highway might be alternatively referred to as a convertible, a car, or a vehicle. In a seminal paper, E. Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976) singled out one level in such hierarchies, which they called the *basic level,* as playing a central role in many cognitive processes associated with categorization. For example, the categories *chair, car,* and *dog* are generally considered to be basic level, and they contrast with more specific *subordinate* concepts (e.g., *recliner, convertible, poodle*) and more general *superordinate* concepts (e.g., *furniture, vehicle, animal*). E. Rosch et al. educed a variety of criteria, each of which converged on a common basic level. For example, the basic level is (a) the preferred in naming, (b) the most abstract level, where members share numerous perceptual properties including overall shape, and (c) the level at which people can categorize most rapidly. It also appears that children learn basic-level categories faster than other levels of categories (e.g., Anglin, 1977; Horton & Markman, 1980; E. Rosch et al., 1976). In short, notions of kindhood may correspond to the basic level.

Of course, observations on the basic level could also be used to buttress similarity-based models of categorization. One could argue (as Rosch did) that objects in the world form natural clusters (at the basic level) and that human cognition is sensitive to these chunks. Presumably, the perceptual system has evolved such that its notion of similarity picks out the basic level as significant (more on this when we consider computational models). A central question, therefore, is whether the basic level changes with development or expertise. If the basic level changes, then its significance may be more in terms of human conceptual processing than facts about the world. Mervis (1987) has argued that the acquisition of new knowledge makes adult-basic-level categories at least somewhat different from child-basic categories, and there is some suggestive evidence for changes with expertise (Mervis et al., 1993; Tanaka & Taylor, 1991). Finally, Mandler, Bauer, and McDonough (1991) have argued that the first categories acquired by children are more global than earlier work had suggested. For example, the first categorization of animals might be into land, air, and sea animals. Mandler (1992, 1993) suggests that global categories reflect the influence of more abstract conceptual notions, such as animate motion (Mandler & McDonough, 1993). So the debate continues.

5. Domain Specificity and Cross-Domain Interactions

Although there is considerable discussion of the domain specificity of conceptual organization in development (e.g., Carey & Gelman, 1991; Hirschfeld & Gelman, 1994; Keil, 1989), space limitations only allow us to do little more than allude to this issue. Most of the research on psychological essentialism has examined biological kinds. One may also ask whether children and adults also believe in a true underlying nature for artifacts, personality traits, occupations, and other sorts of categories. There is evidence that to varying extents they do (Hirschfeld, 1994; Keil, 1989; Rothbart & Taylor, 1992; Yuill, 1992). This raises a variety of important questions concerning cross-domain interactions. Do children start out as essentialists about biological kinds and then generalize by analogy to other domains (e.g., naturalizing social kinds) or do children start out as essentialists about everything and narrow this assumption as domain-specific knowledge is acquired? For differing perspectives on this question see Atran (1998), S. A. Gelman et al. (1994), and Hirschfeld (1995).

6. Summary

Given the significance of learning to theories of categorization, it is easy to see the central role of developmental research in this area. We hesitate to draw any strong conclusions, but it is obvious that key questions about the integration of knowledge (theories) and experience in learning are sharply focused in studies of conceptual development.

D. Cross-Cultural Comparisons

Cross-cultural comparisons represent something of an approach–avoidance conflict for categorization researchers. They may be (logistically) difficult to perform, almost certain to be confounded (in multiple ways) from the point of view of good experimental design, and they often leave the experimenter wondering if a slight difference in procedures might have produced dramatically different results. On the other hand, they provide a powerful test of the generality of some observation or principle. If the same results appear in the face of all the differences between cultures, the results are robust indeed. Furthermore, systematic variation across cultures can likewise be informative. Categorization research has only taken advantage of cross-cultural comparisons in a limited way; nonetheless, the work that has been done has had an important impact.

1. Basic Levels

Rosch's analyses of basic levels reflected a salient influence of cross-cultural comparisons. Berlin, Breedlove, and Raven (1966, 1973) examined folk biological categories across a variety of cultures and argued that there was strong cross-cultural agreement at what one might call the folk generic level. This level more or less corresponds to the genus level of scientific taxonomy. Frequently, the locally represented genus is monospecific, in which case species and genus would be coextensive. This folk generic level of naming biological kinds appears to be a cross-cultural universal. Ethnobiologists have suggested that the basis for this consistency is that organisms possess bundles of correlated features that create natural groupings. Berlin (1992) goes so far as to say that these clusters are "crying out to be named."

The folk generic level may correspond to what E. Rosch et al. (1976) (see also E. H. Rosch, 1975) referred to as the basic level. Indeed the E. Rosch et al. (1976) studies extended the observations about basicness from biological kinds to human artifacts, such as clothing and vehicles. There is, however, a puzzle that remains to be explained. The biological taxonomies that E. Rosch et al. anticipated would be basic level by anthropological (naming) criteria acted like subordinates by the E. Rosch et al. criteria. Rather than *maple, oak, trout, cardinal,* and *eagle* being basic E. Rosch et al. found that *tree, fish,* and *bird* met their criteria for basicness.

Why do the ethnobiological and psychological measures of the basic level disagree? One possibility is that Berkeley undergraduates know little about biological categories relative to the people studied in the anthropological investigations. That is, the basic level may change with expertise (again see Tanaka & Taylor, 1991, for partial support for this idea). A second possibility is that the different measures pick out different levels. Ethnobiological studies tend to use naming or linguistic criteria for basicness, whereas E. Rosch et al. (1976) relied heavily on perceptual criteria. Interestingly, the clearest changes with expertise in the Tanaka and Taylor (1991) studies involved naming preferences. In short, the question of whether the difference is one of expertise or a matter of divergent criteria remains an open one—open and yet central to addressing the question of why we have the categories we have. As we shall see, this question carries over to the next issue.

2. Similarity-Based versus Theory-Driven Learning

Cross-cultural comparisons ought to provide ideal testing grounds for SBL, EBL contrasts. The logic is as follows. People in different cultures have the more or less the same perceptual system (presumably) but differ in knowledge, theories, and beliefs. Therefore, similar categorization systems reinforce the role of SBL, whereas differences support contributions of theory-driven categorization.

But things aren't so simple. First of all, SBL models are sensitive to the distribution of types, and many are sensitive to token frequency. Therefore, differences in categorization may be attributable to differential familiarity or experience with category members. For example, Schwanenflugel and Rey (1986) found that the correlation between typicality ratings for members of common object categories by monolingual English speakers and monolingual Spanish speakers is significantly higher when the contribution of familiarity to judgments is partialed out (see also Boster, 1988). So differences cannot automatically be assigned to the theory side of things. Conversely, similarities do not imply that theory is not in play. People in different cultures may create the same theories (or at least the same kinds of theories). Atran (1987, 1990, 1998) has argued that an essentialist stance toward biological kinds is a cross-cultural universal. These caveats notwithstanding, there have been a number of intriguing cross-cultural comparisons only a tiny sample of which can be presented here (but see Barkow, Cosmides, & Tooby, 1992; Hirschfeld & Gelman, 1994; Premack, 1994; also see Malt, 1995, for a review).

How does one gauge agreement and disagreement across cultures (or even within cultures for that matter)? One very useful tool has been the cultural consensus model (CCM) of Romney, Weller, and Batchelder (1986). The CCM assumes that the agreement between informants is a function of the extent to which each knows the culturally defined consensus or truth. The model assumes a single consensus, that informants' answers are independent, and that each individual can be characterized by a competence parameter that reflects the probability of their knowing the consensus for a given item of information (more knowledgeable people would have a higher competence parameter than less knowledgeable people). If these conditions are met, then a minimum residual factor analysis of the agreement matrix should yield a single-factor solution such that the first latent root should be substantially higher than all other latent roots.

The application of the CCM to categorization is straightforward. If the classification judgment of two distinct groups yields a single consensus (i.e., the CCM model fits the data with a single factor), then a common categorization system is supported. Differences may be observed in at least two ways. First of all the CCM may fail, implying a lack of shared knowledge (e.g., Weller, 1987). Alternatively, the CCM may be generally supported but may not provide a full account of the data. The CCM may be used to calculate the expected agreement between informants, compare it with the observed agreement, and then see if the residual agreement reflects systematic deviations from the consensus. Boster (1986) used this latter technique to evaluate agreement between Aguaruna Jivaro (a South American Indian group) in manioc identification (manioc is a perennial shrub with starchy roots that are an important part of their diet). He observed strong general agreement with the CCM, but also systematic residual deviations from consensus that were correlated with the kinship distance of the participants.

Much of the anthropological work on culture and categorization has focused on biological kinds. On a general level, there is a striking amount of cross-cultural consistency in categorization (e.g., Boster, 1987; Boster & D'Andrade, 1989; Malt,

1995) and substantial agreement of folk biology with scientific taxonomy (e.g., Atran, 1998; Boster, Berlin, & O'Neill, 1986). Interestingly, in their studies of similarity comparison by fish experts (peoples who fished for a living) and novices, Boster and Johnson (1989) found that expertise was associated with *decreased* agreement with scientific taxonomy. Boster and Johnson suggest that knowledge of function influenced the judgments of experts and led to at least some category reorganization.

The categorization function is only one aspect of conceptual behavior. Unfortunately, there has been very little work on the use of categories in reasoning. Walker (1992) examined preservation-of-identity judgments by rural, urban poor, and urban wealthy Nigerian adults. Participants heard stories describing changes where one natural kind came to appear like another (as in the studies of Keil and of Rips described earlier). Furthermore, these changes were described as taking place either in a ritual context or a nonritual context. Walker found that the Nigerian participants preserved identity of category membership essentially all the time in nonritual contexts and the vast majority of the time in ritual contexts. She also observed that nonpreservation judgments varied as a function of the centrality of the concept to ritual practices (dogs are more central than chickens, and there were more nonpreservation judgments involving dogs than chickens). These latter findings suggest that adults' natural kind concepts may be at least partially influenced by belief systems other than the biological (see also Boyer, 1990, for further discussions of the interplay between biological reasoning and religious/cultural beliefs).

Atran (1998) has recently examined relationships between categorization and reasoning, using both the CCM and the Osherson, Smith, Wilkie, Lopez, and Shafir (1990) category-based induction model. His preliminary results suggest a close correspondence between folk biological taxonomy and hypothetical reasoning about reproduction among the Itza Maya of Guatemala. For example, Atran finds both similarity and typicality effects in inductive reasoning. It also appears, however, that the basis for typicality may deviate from central tendencies for the Maya. For example, the most typical bird is the *turkey,* priced for its meat and culturally significant, and the most typical snake is the *fer-de-lance,* the most poisonous of snakes. In short, for some categories typicality may be driven by proximity to an ideal rather than an average (as in Barsalou's, 1985, observations with goal-derived categories). The fact that typicality effects are observed on the induction task suggests that the cultural differences are not based on misunderstandings about the meaning of "typicality." Given that most college students are not especially knowledgeable about biological kinds, it is possible that the cultural differences in the basis for typicality (ideals versus central tendencies) should be attributed to differences in knowledge and expertise.

3. Summary

We end this section on an optimistic note. It appears that cross-cultural similarities in categorization are strong enough to avoid the confusion that might have been created by a morass of differences. This backdrop of agreement permits focused questions about differences, and the CCM and category-based induction model provide important methodological tools for these analyses. Cross-cultural comparisons are not a panacea for working out relationships between similarity and theory, but they do represent a fruitful avenue for exploration.

E. Computational

1. Issues and Purposes of Modeling

In research on categorization, there is a widespread tradition of implementing theoretical ideas as computational or mathematical models. This development of models of categorization has had several purposes. Foremost, a categorization model is a precise statement of an account of categorization. Modeling may be thought of as a language for describing theories of categorization. Stating differing accounts of categorization within this common language makes it easier to compare them. As previously mentioned, Nosofsky (1991) showed that when Anderson's rational theory (1990) and Medin's context theory (Medin & Schaffer, 1978) are compared, the context theory is equivalent to a special case of the rational theory. Also, describing categorization with the language of modeling has an advantage over, say, the language of English. Models run.

Modeling provides some insurance against potential errors of human reasoning; it is often difficult for a researcher to know what some theory will predict until the theory is implemented as a model (Hintzman, 1991). For example, Medin and Schaffer (1978) discovered that an exemplar model of categorization can predict prototypicality effects; that is, new prototypical category examples may be categorized more accurately than old, less central category members. This conclusion was surprising because exemplar models do not assume that a category prototype is explicitly represented in memory. Therefore, findings of prototypicality effects in experimental data need not be interpreted as indicating that people form prototypes.

It is important to note that no computational model (so far) has been presented as a complete account of categorization. For example, many models assume that category members are represented in terms of lists of features, but these models do not provide an account of how people learn to use these features. Sometimes it is a virtue of modeling that it allows us to focus on critical aspects of theory, while keeping other issues in the background. To give an extreme example, Busemeyer, Myung, and McDaniel (1993) derived predictions for a set of connectionist models of category learning, which are largely independent of how features are represented and of the specific algorithm for learning. Busemeyer et al. showed that none of these models can account for a phenomenon in human category learning known as the *cue competition effect,* in which learning about a valid cue is overshadowed by knowledge of another valid cue. What is critical about these models is that they all learn optimal associations between cues (or features) and categories, but the cue competition effect results from suboptimal learning. Busemeyer et al. suggested modifications for these models that would allow them to better approximate human learning.

A final general point we will make is that categorization models, ideally, will not be isolated accounts of a particular task or experiment but instead will dovetail with other theoretical accounts of cognition. We believe that categorization is an important topic, but it is a topic that is intertwined with the study of learning, memory, and reasoning (among other topics). One example of the potential synergy between categorization models and other computational models of cognition is the compatibility between exemplar models of categorization and multiple-trace models of memory (Gillund & Shiffrin, 1984; Hintzman, 1986, 1988). Multiple-trace models assume that a memory judgment, such as a recognition decision, depends on evaluating the total similarity of a test item to memory traces of particular stimuli (see C. M. Jones & Heit, 1993, for a review). Likewise, exemplar models assume that a decision whether to place a test item in one category or another depends on evaluating the similarity of the test item to memory traces for members of each category. Much research has capitalized on this connection between categorization and memory, and has led to the development of unified accounts of not only categorization but also other memory abilities, such as recognition, frequency judgment, and recall (Estes, 1993, 1994; Heit, 1993, 1994; Hintzman, 1986, 1988). Note that such a synergy between models of different, related tasks need not be limited to the common framework of exemplar models and multiple-trace models. For example, connectionist modeling provides another framework for developing general models of categorization and other cognitive abilities.

2. Learning

Now it is time to discuss and compare particular models of categorization. Most of these models fit into one of two broad categories: SBL or EBL. Briefly, what SBL models have in common is the assumption that a judgment about how to categorize something depends on its similarity to previously observed category members. The SBL models differ mainly in how the information about past category members is represented in memory (e.g., in terms of summary statistics in an abstraction model or in terms of association strengths in a neural network model). In contrast, the fundamental criterion for categorization in EBL models is that a category must provide an explanation, just as a theory provides an explanation for observed data. To use a well-known example from Murphy and Medin (1985), a fully clothed person who jumps in a swimming pool may be categorized as a fraternity pledge, because *fraternity pledge* provides an explanation for this person's odd behavior. In this example, it seems difficult to explain the categorization of this person in terms of similarity to other pledges.

a. Similarity-Based Learning

i. Abstraction models One attractive idea about how to represent concepts is to store an abstraction, that is, summary information about category members. For example, to categorize cows, you might keep in memory what an average cow looks like. Most cows that you will see probably resemble this cow prototype (e.g., in terms of shape overlap). The parsimony of this approach is appealing. Indeed, some computer schemes for object recognition (Ullman, 1989) rely on templates to represent average category members. In addition, prototype models are important historically in cognitive psychology. Posner and Keele (1968) found that subjects, after viewing members of a category of dot patterns, were quite likely to say that the prototype (i.e., the central tendency of the category) was also an observed category member even when the prototype had not been presented. This result is certainly suggestive of the claim that people form abstractions of central tendencies. (For a review of successful applications of prototype models, see Hampton, 1993.) However, psychological research has progressed beyond pure prototype models, mainly because it is known that people learn more about categories than just their central tendencies.

In addition to the central tendency of a category, people can learn about the variability of a category. For example, Fried and Holyoak (1984) taught subjects about two categories of painting where for each category, the members were distortions of the category's central prototype. However, one category of paintings had quite variable members, and the members of the other category were close to its prototype. In the critical test questions, subjects were asked to categorize a new painting that was midway between the prototypes of the two categories. Most people placed the new painting in the more variable category. If subjects had simply remembered the central tendency of each test item, then they would have been indifferent between the two categories. Fried and Holyoak concluded that an abstraction model that only stored central tendencies was untenable. They suggested that abstraction models could be improved so that people would learn about variability as well (e.g., they learn summary information about means and variances for each dimension; see also Flannagan, Fried, & Holyoak, 1986).

Much more elaborate abstraction models have also been developed. According to general recognition theory (Ashby & Gott, 1988), a perceptual category is represented in memory by a specification of a multidimensional probability density function. The category description is probabilistic because the same category member may be processed in somewhat different ways on different observations, due to error or noise in the perceptual system. The representational system of general recognition theory is powerful; a category description may contain information about expected means and variances of various stimulus dimensions as well as correlations between dimensions. In addition to these representational assumptions, the framework of general recognition theory allows a variety of processing assumptions about how people make categorization decisions. In one version of the theory, a person sets a decision boundary, such as a line or curve, between two categories. For observations that fall exactly on this boundary, it is equally likely that the observation comes from each of the two multidimensional distributions corresponding to the two categories. However, most stimuli will fall on one side of the categorization boundary or the other, so their category membership will be clear.

ii. Exemplar models The modeling framework that seems diametrically opposed to abstraction models is exemplar models. Rather than representing a concept as a summary of what the category members are like, in exemplar models the concept is represented by memories of particular category members. Note that exemplar models of categorization by humans do not necessarily assume that all category members are remembered, or that each category member is stored veridically. We know that human memory is not perfect. Still, there are reasons to suggest that people represent categories not simply with abstract summary information but with information about specific instances.

Some of the best arguments for the plausibility of exemplar representation have been provided by Brooks (1978). First, he suggests that some categories may not have an obvious abstract representation, so remembering exemplars of the category would be an attractive alternative. Likewise, other factors such as time pressure during learning might make deriving a summary difficult. Second, category learning may occur in the pursuit of other goals that would require learning about particular instances. For example, when we learn about social categories, such as categories of people with various occupations, we are often interested in learning about these people as individuals as well. It is plausible that this information about individuals could be accessed during judgments about categories. Certain learning conditions, such as repeated experiences with the same individual, would also facilitate exemplar memory. Finally, remembering exemplars allows for maximal flexibility for conceptual organization in the future. We might need to form different concepts on different occasions or for different goals (see Barsalou, 1983). For example, information in memory about a particular pet dog could later be used for judgments about the categories *pet* and *dog*. (See Heit, 1992, for an application of an exemplar model to studies in which subjects first learned exemplars and later made multiple categorization judgments.)

One widely tested exemplar model of categorization is the *context model* (Medin & Schaffer, 1978). Consider the task of deciding whether some stimulus, x, belongs in category A or category B. According to the context model, the probability of categorizing x as an A rises with the similarity of x to members of A and falls with the similarity of x to the members of B. In the context model, stimuli are represented as vectors of features. For example, in a study of simulated medical classification by Medin, Altom, Edelson, and Freko (1982), it was assumed that exemplars of categories of people with a certain disease were represented by a list of symptoms, such as *swollen eyelids* and *discolored gums*. The similarity between two stimuli is assessed by counting their numbers of matching and mismatching features. In general, similarity is assumed to be monotonically related to the number of matches, but the critical assumption of the context model, known as the *multiplicative similarity rule,* is that perfectly matching stimuli or near-perfect matches will be considered much more similar than pairs that mismatch on several features. Thus, in the categorization of x, exemplars that are close matches to x will especially influence categorization.

One implication of the high impact of close matches is that the context model predicts that people can readily learn categories that are not *linearly separable.* That is, the distributions of the members of two categories may overlap, so that it would be impossible to draw a straight-line boundary that segregates the members of one category from the other. The context model predicts that people can learn to distinguish between such categories because a new stimulus that is a close match to an old exemplar will likely be placed in the same category as that exemplar, whether or not the two categories overlap. In contrast, abstraction models such as prototype models or simple linear classifiers (both special cases of the Ashby & Gott, 1988, framework) predict that people cannot learn to distinguish correctly between categories that are not linearly separable. It is clear that people can learn nonlinearly separable categories (Medin & Schwanenflugel, 1981), although more recent evidence (e.g., Wattenmaker, 1995) suggests that people's relative ability to learn linearly separable versus nonlinearly separable categories may vary across domains.

One limitation of the context model is its simple representation of stimuli as vectors of features. Consider learning about the people you meet, including their physical appearances and their intellectual abilities. It would seem valuable to represent continuous information such as height in inches and probably IQ, rather than just features such as *short* and *intelligent.* Nosofsky (1986) has proposed a *generalized context model* (GCM) that is more flexible in how it represents stimuli and in how similarity is evaluated. According to the GCM, stimuli are represented as points in multidimensional space. For example, a short intelligent person might be represented with a point near 50″ on a height axis and near 140 on IQ axis. As in the context model, in the GCM, categorization decisions are made by comparing a stimulus to exemplars retrieved from memory. Nosofsky has evaluated several rules for assessing the similarity between stimuli described in terms of multiple continuous dimensions. For example, one consideration is whether the dimensions are separable (as in the case of height and intelligence) or perceptually integral (such as hue and saturation of colored objects, see Garner, 1974) (Nosofsky, 1987). Together, the context model and the GCM has been applied successfully to the results of many laboratory studies on categorization (see Nosofsky, 1988, 1992, for reviews).

To complete the discussion of exemplar models, we will briefly mention two additional applications of exemplar theory. First, recent work by Eliot Smith (E. R. Smith, 1990; E. R. Smith & Zarate, 1992) has applied exemplar models to the domain of social cognition. This research has demonstrated with computer simulations that many phenomena in social psychology, such as the influence of knowledge about social categories (i.e., stereotypes) on judgments about individuals persons, can be explained in terms of exemplar models. Furthermore, people evaluate the similarity between individuals in a somewhat flexible manner. In some contexts, making a judgment about some person might lead to retrieval of memories of other persons of the same gender, and in other contexts, memories of persons of the same racial group might be retrieved instead. However, the issue of flexibility of similarity is not an entirely solved problem; it is still largely an open question how people's learning and reasoning processes lead them to focus on certain dimensions on certain occasions.

Second, exemplar theory also provides the basis for a successful technique in AI models, case-based reasoning (Bareiss, 1989; Hammond, 1989; Riesbeck & Schank, 1989). A case-based reasoning program can make inferences about a new case by retrieving similar cases from memory. For example, the computer program MEDIATOR attempts to solve international conflicts by retrieving case studies of similar past conflicts and determining a solution from the past conflicts and how they were resolved (Kolodner & Simpson, 1989). Comparing instances also provides the opportunity for developing representations intermediate between exemplars and fully abstract representations (e.g., Medin & Edelson, 1988; Ross & Kennedy, 1990; Ross, Perkins, & Tenpenny, 1990; Spalding & Ross, 1994). One challenge for case-based reasoning models is known as the *indexing problem*. How shall memory be organized so that in reasoning about some new situation, the most helpful or relevant past cases will be considered? The indexing problem is a version of the problem faced by all similarity-based models of categorization. If categorization depends on similarity to a representation of known category members, then what features or dimensions of the representation are counted in evaluating similarity (Murphy & Medin, 1985)?

iii. Connectionist models An increasingly popular framework for developing computational models is *connectionist,* or *neural network,* models (McClelland & Rumelhart, 1986; Rumelhart & McClelland, 1986). Connectionist models of categorization usually consist of a set of input nodes corresponding to features on stimuli to be categorized, a set of output nodes corresponding to possible response categories, and a network of connections between the input and output nodes. This network of connections stores associative information about the relations between inputs and output (i.e., between stimulus features and categories). These connections may have a simple structure, such as one connection between each input node and each output node, or the structure may be more complex, with multiple layers of connections as well as additional nodes, known as hidden traits. Learning takes place by adjusting the connection strengths according to a learning rule (much current research on connectionist modeling addresses the development of learning rules). Typically, the degree of learning about a stimulus is proportional to a measure of categorization error. If the connection strengths already allow a particular stimulus to be categorized perfectly, then the connections would not be adjusted after an observation of this stimulus.

One particular application of connectionist principles to categorization is Kruschke's (1992) ALCOVE model. As in other connectionist models, ALCOVE has input nodes corresponding to the dimensions of variation of stimuli and output nodes corresponding to response categories. In addition, ALCOVE has two kinds of connections. First, attentional connections run from the input nodes to a set of hidden units that correspond to representations of specific exemplars in memory. For example, if the input nodes corresponding to *small* and *red* are activated, then exemplars of hidden units corresponding to small, red things will be especially activated. In this way, ALCOVE is an exemplar-based connectionist model, a hybrid between neural nets and traditional exemplar models. Second, connections run from the hidden units to the output nodes. These connections are adjusted as in other connectionist models so that ALCOVE can learn correct categorizations.

The novel contribution of ALCOVE lies in the first kind of connections, which corresponds to attentional weights on different dimensions. For example, if the attentional connections for color input nodes are stronger than connections for size input nodes, then an input of *small and red* would tend to activate *red* units more than *small* units. Because ALCOVE is a connectionist network, these attentional strengths are learned by the model. Previous work on attention in category learning (e.g., Nosofsky, 1986) has shown that people selectively attend to diagnostic dimensions, but previous models have not provided an account of the process by which selective attention is applied. In contrast, ALCOVE provides an account of how people learn to attend to the dimensions of stimuli that are most useful or diagnostic for learning categories. For some examples of successful applications of ALCOVE to studies of categorization, see Kruschke (1992), Nosofsky and Kruschke (1992), and Nosofsky, Kruschke, and McKinley (1992).

An alternative connectionist model for category learning has been proposed by Gluck and Bower (1988b). In this model, input nodes are connected directly to the output nodes, without any hidden units. One advantage of this relatively simple scheme is that it makes it possible to derive the predictions of the model without running extensive simulations. For example, Markman (1989) derived mathematical predictions of how sensitive this model will be to base rates (i.e., the relative frequencies of categories). Although Gluck and Bower's model does not have the hidden units, or the separate component for attentional learning of the ALCOVE model, it has still been quite successful (see Gluck & Bower, 1988a, 1988b, for examples of applying this model to particular studies). One problem with the original model is that it can only learn linearly separable categories (recall that Medin and Schwanenflugel [1981] found that people can learn categories whether or not they are linearly separable). This limitation of the model follows directly from its lack of hidden units; certain patterns of categorization require hidden units (Minsky & Papert, 1988). Even with hidden units many network models predict that linearly separable categories should be easier to learn than nonlinearly separable categories, a prediction for which there is little or no support. More recently, Gluck and Bower (1990) have proposed an extension known as the *configural cue model*. In the configural cue model, input nodes may correspond not only to features of stimuli but also to pairs of features, triples of features, and in most extreme case, all possible *n*-tuples of features. In this extreme case, the input to the model would correspond to an exemplar representation, because each possible stimulus would have its own input node (see also Barsalou, 1990).

iv. Rule-based models Rule-based approaches to categories and concepts have had an uneven history. They received a lot of attention in the 1960s and early 1970s when experimenters implicitly or explicitly acted as if categories are well defined. Interest was in the relative difficulty of different forms of logical rules (e.g., Bourne, 1970). If categories are fuzzy or ill defined, then most such logical rules will not work. Consequently, one by-product of the shift to the probabilistic view was severely diminished attention to rule-based models.

In principle, however, fuzzy structures do not rule out rules. Indeed many participants in laboratory studies of categorization report that they are trying to develop rules and they often succeed (by forming disjunctive rules or rules with exceptions). Therefore, ignoring rule learning may come at some risk (Martin & Caramazza, 1980; Medin, 1986). Nosofsky, Palmeri, and McKinley (1994) have shown that a rule-based model can account for a wide range of categorization phenomena associated with fuzzy artificial categories, so this view needs to be taken seriously.

Any one category partitioning is consistent with a virtually limitless set of rules or inductive generalizations. Michalski (1983a, 1983b) has developed an AI system for inducing rules from partitioned examples where the aim is to have rules that are psychologically comprehensible. Michalski's INDUCE system works by selecting an example and describing it in alternative ways according to certain generalization rules. The goal of these generalization rules is that the description will apply to other examples of the same category and not to examples of contrasting categories. The process is recursive such that if the best rule is consistent (does not have any counter-examples) but not complete (fails to apply to all category), an example that is not covered by the rule is selected and generalization rules applied to it. Therefore, the general form of rules will be a disjunction of conjunctions. Michalski has found that INDUCE shows very good accuracy in classifying new examples after inducing a rule from a modest set of training examples (e.g., Michalski, Mozetic, Hong, & Lavrac, 1986; also see Michalski, 1993, for more recent applications of the same general framework).

Medin et al. (1987) compared the rules developed by INDUCE with those of people, again using preclassified examples. They observed both general agreement and systematic differences. In particular, people often develop initial rules that are overly general and then restrict them by adding clauses that eliminate counter-examples (e.g., a rule such as "X and Y" that applies to some examples of the alternative category may be patched up by adding the hedge "and not Z"). Medin et al. suggested that this rule formation strategy facilitates the inference function of concepts.

There are other similarity-based AI classification systems that are conceptually close to rule-based models. For example, discrimination net models that involve a series of branching tests (of feature values) can be construed as rule generators (see also Rendell & Cho, 1990). Increasingly, these models are being applied to human categorization data (e.g., Ahn & Medin, 1992; Richman, 1991; and especially Fisher & Langley, 1990, and Fisher & Yoo, 1993). A good recent review of relevant work in this area is provided by the volume edited by Fisher, Pazzani, and Langley (1991).

v. Comparing and developing models For the most part, we have emphasized the development of the similarity-based models within single frameworks, such as the progression from the prototype model of Posner and Keele (1968) to the abstraction models of Ashby and Gott (1988), the progression from the exemplar model of Medin and Schaffer (1978) to the exemplar model of Nosofsky (1986), and the progression from the connectionist model of Gluck and Bower (1988b) to the configural cue version of this model (1990), rather than emphasizing comparisons between frameworks, such as whether exemplar models are better than abstraction models. This emphasis reflects a bias of our own, that computational modeling is a language that is particularly useful for expressing and developing theories of categorization. To continue the language metaphor, different kinds of modeling, such as connectionist models and exemplar models, may be thought of as different languages. It is easier to compare two models within the same framework than two models in different frameworks, just as it is easier to compare two short stories in one language than two stories in different languages. Certainly it is also valuable to compare models from different frameworks; for example, Nosofsky (1992) has shown that certain abstraction, exemplar, and connectionist models are formally equivalent. One implication of these analyses is that you can't tell a model by its label—an exemplar model and a connectionist model may be much more similar than two connectionist or two exemplar models. Furthermore, a complete description of a particular model must refer not only to its form of representation but also must develop its processing assumptions (Barsalou, 1990).

b. Explanation-Based Learning

i. Arguments for theory effects What the similarity-based models (abstraction, exemplar, connectionist, and rule-based) described so far have in common is that they form categories with bottom-up, data-driven processes. Murphy and Medin (1985) have argued that such accounts of category learning are incomplete, because the concepts that people form cannot be predicted only from what people observe (see also Schank et al., 1986). People bring to bear many forms of prior knowledge that also influence concept formation, from simple expectations about what will be in a category and what features will be relevant for evaluating similarity, to more elaborate causal knowledge (or theories) about the relations between category members and the relations between categories. Many recent studies have demonstrated the influence of prior knowledge on category learning (e.g., Hayes & Taplin, 1992; Heit, 1994; S. S. Jones, Smith, & Landau, 1991; Lamberts, 1994; Murphy & Wisniewski, 1989; Pazzani, 1991; Wattenmaker et al., 1986; Wisniewski & Medin, 1991, 1994a; see Murphy, 1993, for a review). For example, Wattenmaker et al.

(1986) showed that prior knowledge of occupations helped subjects learn about novel categories of occupations. Together, these studies have made it clear that it is easier to learn new categories that are consistent with prior knowledge than categories that are inconsistent with prior knowledge. In addition, people's beliefs about newly formed categories reflect knowledge from outside of these categories; that is, people show assimilation effects.

ii. AI models of EBL A number of proposals for addressing the role of prior knowledge and theories have been developed in the domain of machine learning under the banner of EBL (Dejong, 1988; Ellman, 1989; Mooney, 1993). Typically, an explanation-based system uses its background knowledge (in the form of a domain theory) to explain or prove why a training example is a member of a given category. It then generalizes the explanation so that it will apply to future examples. An advantage over SBL approaches is that appropriate abstraction may take place on the basis of experience with only a single example.

There have been some promising recent applications of EBL to psychological experiments (e.g., Ahn, Brewer, & Mooney, 1992; Pazzani, 1991). AI research in the EBL framework has been very active, addressing a variety of issues, such as the problem of incomplete or incorrect theories (e.g., Porter, Bareiss, & Holte, 1990; Rajamoney, 1990). These and other analyses (e.g., Dietterich, 1986) have led to a growing interest in methods for integrating theory and data (EBL and SBL) a topic to which we now turn.

iii. Merging SBL and EBL An alternative to these pure EBL accounts is to develop existing SBL models to address the effects of prior knowledge, leading to what may be considered mixed models. As one example, we describe in some detail a recent similarity-based model of category learning that has been modified to address the effects of prior knowledge (Heit, 1994). The *integration* model is an exemplar model that is a variant of context theory (Medin & Schaffer, 1978). The novel assumption of the integration model is that two kinds of exemplars influence judgment of whether some stimulus belongs in a category: both exemplars of that category as well as *prior examples,* from other categories. For example, imagine that you move to a new city and you are looking for friends to join you in jogging. In effect, you are trying to learn about a new category: joggers in this city. In your early observations of residents of the city, your categorization judgments about whether these persons are joggers will be influenced by similarity to prior examples of joggers from other contexts (e.g., joggers from where you previously lived). Eventually, with more observations, your categorization judgments will be influenced much more by observed examples of joggers in your new city and less by the prior examples. In this way, the integration model is similar to Bayesian models of statistical estimation (Edwards, Lindman, & Savage, 1963). For several experiments simulating this experience of category learning in a new context where subjects were knowledgeable about prior examples, Heit (1994) found that the integration model gave a good qualitative and quantitative account.

In addition to the integration of prior examples and observed examples, Heit distinguished other possible processes by which prior knowledge might affect category learning. First, prior knowledge may lead to selective weighting of category members so observations that fit prior knowledge are remembered best. For example, you might be more successful at learning about joggers who own expensive running shoes than about joggers who do not own expensive running shoes. Second, prior knowledge may lead to selective weighting of features. In learning to categorize joggers in your new city, you might attend to people's shoes rather than hair color. Third, prior knowledge may have a distortion effect; for example, a jogger without expensive running shoes might be misremembered as a jogger with expensive running shoes or even as a nonjogger. Although these additional processes all seem plausible, the results of Heit (1994) could be explained without any of them (i.e., by the integration model alone). Thus, providing distinctive evidence for when these other processes occur in addition to integration will be a task for future research.

Of course, there are a number of other ways of integrating EBL and SBL. For example, Lebowitz describes a system, UNIMEM, where SBL is used to determine regularities, and then UNIMEM attempts to explain these commonalities with its domain theory. The goal of the explanatory component is to separate causally relevant features from those that are spurious or coincidental. Other AI systems such as Induction over the Unexplained (IOU) (Mooney, 1993), or Induction over the Explained (IOE) (Flann and Dietterich, 1989), and EXOR (Fisher & Yoo, 1993; Yoo & Fisher, 1991) operate in essentially the opposite manner. For example, IOU first applies EBL on training items, and features that do not enter into explanations are input to an SBL component. The target concept is then augmented with these unexplained regularities. The SBL component allows the system to acquire predictive features that are not covered by the domain component. IOE develops explanations for each of a series of training examples and then employs empirical learning (SBL) to detect frequently occurring substructures and patterns of features across explanatory trees. For a review and analysis of integrated systems, see Mooney (1993) and Wisniewski and Medin (1991).

A generalization that applies to almost all integrated SBL–EBL systems is that the interaction between the empirical and explanatory components is unidirectional and indirect. In a number of systems the first component acts as a filter for the second component by reducing the number of features input to the second component. Wisniewski and Medin (1991, 1994a, 1994b) have argued, however, that more tightly coupled systems are needed, at least in the case of psychological process models. They used categorization and rule induction paradigms where the same examples (children's drawings) were associated with different domain theories (e.g., in one case people might be told the drawings were done by creative versus noncreative children; in another that the drawings were done by emotionally disturbed versus mentally healthy children). A number of findings point to the need for greater interaction between theory and data. First of all, it may not always be reasonable to assume a space of prespecified unambiguous features. Wisniewski and

Medin noted that the features comprising subjects' rules varied as a function of domain theory and that the same aspect of a drawing was interpreted differently for different category labels. Participants also sometimes reinterpret features when given feedback about category membership. Finally, participants' rules often involve abstract features that are operationalized differently as a function of learning history (e.g., one might expect creative children to draw detailed pictures, but how detailed does a drawing have to be to qualify as "detailed"?). On the basis of these observations, Wisniewski and Medin argued that relatively modular ways of incorporating prior knowledge into categorization models are inadequate.

3. Induction

The extent of the previous section reflects that most work on computational models of concept use has focused on one conceptual function, namely categorization. However, other work on computational models has addressed the inference and inductive reasoning function of concepts. The category-based induction (CBI) model (Osherson et al., 1990; Osherson, Stern, Wilkie, Stob, & Smith, 1991) addresses the issue of how we infer novel properties of categories. For example, given the premise that cows have sesamoid bones, how likely is the conclusion that dogs also have sesamoid bones? According to the CBI model, two factors influence how people evaluate the inductive soundness of such inferences. First, the similarity between the premise category (e.g., *cow*) and the conclusion category (e.g., *dog*) is critical. To the extent that people believe that the premise and conclusion categories share other properties, people will be willing to project a new property from the premise to the conclusion. So if the premise is that elephants (instead of cows) have sesamoid bones, it would seem less likely that dogs have sesamoid bones, because dogs and cows are more similar than dogs and elephants.

The second factor in the CBI model is the *coverage* of the premise, that is the similarity between the category or categories in the premise and members of the superordinate category that encompasses the categories in the premise and conclusion. A few examples should make coverage clear. Consider again an inductive inference from *cow* to *dog*. The most specific superordinate category that includes cows and dogs is *mammal*. Now, *cow* is fairly similar to other members of the category *mammal;* cows are moderately typical mammals. Thus, if cows have sesamoid bones, it is moderately plausible that all mammals have sesamoid bones. In turn, if all mammals have this property, than so must dogs have the property. In the CBI model, the two sources of evaluating inferences, similarity and coverage, are just added together. Category members that are atypical do not contribute much to coverage; for example, *aardvark* would provide little coverage for the superordinate category *mammal* (see also Rips, 1975). The CBI model also provides an elegant way to evaluate the coverage of arguments with multiple premises. For example, given the premises that both horses and squirrels have sesamoid bones, it seems likely that all mammals have sesamoid bones, because horses and squirrels are quite diverse members of the superordinate, *mammals.* On the other hand, the premises that horses and mules have some property does not lend as much support to the belief that all mammals have the property, because horses and mules do not cover the superordinate category *mammals* much better than just horses alone.

The CBI model provides an intuitively pleasing account of how people use similarity and category information to make inferences, and gives a good explanation for many empirical results at both a qualitative and quantitative level (see Osherson et al., 1990, 1991). The CBI model focuses on the role of categories in inductive reasoning, but it does not provide an account of the role of properties. Most of the Osherson et al. (1990, 1991) studies used "blank properties" such as *has sesamoid bones* and *has an ulnar artery* that seem biologically related but are otherwise unfamiliar. Heit and Rubinstein (1994) have proposed that one systematic effect of the property being inferred is that it leads people to focus on certain other relevant properties when evaluating similarity. For example, Heit and Rubinstein found that for the anatomical properties such as *has a liver with two chambers,* subjects were more willing to make an inference from *mouse* to *bat* than from *sparrow* to *bat*. Thus, for an anatomical property, people are influenced by the common biological properties of *mouse* and *bat.* In contrast, when subjects evaluated inferences about behavioral properties, they appeared to evaluate similarity in terms of other behavioral characteristics, such as method of locomotion. For example, for the behavioral property *travels shorter distances in extreme heat,* subjects favored inferences between *sparrow* and *bat* over inferences between *mouse* and *bat.* Presently, the CBI model does not address the part of inductive reasoning by which we determine which properties are relevant to evaluating similarity, nor does any other model address this issue.

4. Conceptual Combination

Finally, we address the computational modeling of another important function of knowledge of categories, conceptual combination. Conceptual combination is important because it allows us to be productive in our use of concepts, so that we can create and understand a potentially unlimited set of new concepts. For example, we are only rarely faced with the task of understanding a single concept in isolation; much more commonly we need to interpret a set of concepts combined to form a phrase or sentence. Yet most categorization models do not address the phenomena related to combined concepts (Rips, 1995). One exception is the selective modification model (E. E. Smith & Osherson, 1984; E. E. Smith, Osherson, Rips, & Keane, 1988), which is a specialized account of the understanding of adjective-noun combinations such as brown apple. According to the selective modification model, to understand this term a person would retrieve the prototype of apple, pay extra attention to the dimension of color, and replace the default value of red for apples with the value brown. In effect, the person would be constructing a new prototype for brown apple by modifying the apple prototype. Then, judgments such as categorization decisions and typicality ratings could use the brown apple prototype in accordance with an ordinary prototype model of categorization.

But conceptual combination is more complex than the selective modification model (or any other current model) suggests. One complication is that the people also use their general knowledge about relations between features. For example, in interpreting the combined concept *large spoon*, people not only modify their spoon prototype to make it larger, but they also seem to make inferences about other features, such as that a large spoon is likely to be made of wood (Medin & Shoben, 1988). Interestingly, conceptual combination sometimes leads to inferences about emergent features (i.e., features of the combined concept that are not considered as features of either constituent concept) (Hampton, 1987; Hastie, Schroeder, & Weber, 1990; Kunda, Miller, & Claire, 1990; Murphy, 1988; Rips, 1995). For example, many people would expect a Harvard-educated carpenter to be a nonconformist, but this feature is not as often expected for either of the constituent categories, Harvard-educated people and carpenters (Kunda et al., 1990). These two findings, regarding feature relations and regarding emergent features, indicate that conceptual combination could be a compositional process, as assumed by the selective modification model (as well as the model of Hampton, 1987, 1988) but that other sources of knowledge influence understanding. In other words, you cannot fully interpret a combination of two concepts by simply combining what you know about each concept alone. Instead, additional knowledge outside of the two constituent concepts must also be brought to bear (Hampton, 1988; Kunda et al., 1990; Murphy, 1988; Rips, in press).

5. Summary

Our summary paragraphs in this review may be seen as not only modular but interchangeable. Again we see the tension between SBL and theory-based category learning and a number of attempts to reconcile the two. We don't think it is speculative to predict that the integration of SBL and EBL will be a recurrent theme in future research (see the edited volume by Nakamura, Taraban, & Medin, 1993, for a number of examples).

F. Neuroscience

Opportunities to link categorization research with observations from neuroscience have only recently begun to be exploited. In this section we provide only a single example, but one that shows the potential interplay of neuroscience with categorization.

It appears that brain damage can lead to category-specific impairments of semantic memory. Warrington and Shallice (1984) reported that some patients were much worse at identifying plants and animals than nonliving things (see also Sartori & Job, 1988). Impaired knowledge of nonliving things relative to living things has also been observed (Warrington & McCarthy, 1987). These observations suggest that knowledge of living kinds versus inanimate objects are represented in distinct submodules of semantic memory. Warrington and her associates (e.g., Warrington & McCarthy, 1983) have raised the alternative possibility that semantic memory is organized in terms of modality. They note that living kinds are distinguished primarily by their perceptual sensory features, whereas artifacts are distinguished in terms of functional properties (e.g., intended use). In brief, the selective impairment of knowledge about living versus nonliving things may be based on an underlying perceptual versus functional information loss.

Farah and McClelland (1991) have recently described support for Warrington's view in the form of computational modeling. They developed a parallel distributed processing model organized by modality (perceptual vs. functional) and demonstrated that simulated lesions to either the functional or visual area of the network could reproduce impaired knowledge of living things versus artifacts. Furthermore, the simulations provided a detailed account of more subtle aspects of these deficits (e.g., differences between probing knowledge visually vs. verbally).

The neuroscience observations and the explorations with artificial neural networks operate hand in hand. Neuroscience provides hypotheses to be investigated as well as constraints on models. The models, in turn, allow one to go beyond a macroscopic level of analysis, and their explanatory power can feedback to suggest further relevant observations. We anticipate that categorization research increasingly will benefit from these sorts of interdisciplinary analyses.

III. CHALLENGES AND OPPORTUNITIES

A. Challenges

It is perhaps a sign of our optimism that we would not be unhappy with a straight line projection of current states of affairs in categorization research. By projection we mean continuing to follow promising paths, not exactly more of the same. Our sanguine attitude does, however, recognize certain significant challenges.

1. Questions of Features

Although it is often advantageous to use stimulus materials where the constituents or features are readily described, little progress has been made in understanding how features come into existence. In the case of certain visual features one might claim that they are hardwired into the perceptual system, but this sort of analysis is unlikely to work for more conceptual properties. There is some work on feature construction in neural network models (e.g., Rumelhart & Zipser, 1985), but so far there has been little by way of application to categorization. We need to understand how features come into play and how they are developed and modified as a function of knowledge structures and experience (see Wisniewski and Medin, 1994b, for more discussion).

2. Question of Kinds

Any research program that focuses on a narrow range of stimulus materials or tasks faces the twin dangers of narrow generality or misattribution. With respect to the latter, the combination of a single task and stimulus type creates an ambiguity as to whether the performance observed is triggered by the task, by the stimuli, or by some interaction of the two. Lee Brooks and his associates (e.g., Brooks et al., 1991; Reagher & Brooks, 1993; Whittlesea, 1933) have repeatedly demonstrated powerful interactions of stimuli with processing tasks. More generally, categorization research has tended to focus primarily on nominal, natural kind, and artifact categories. It is not clear how or how much our perspective might change if there were corresponding efforts to study actions, events, or any of a variety of other types of categories. For example, Gentner (1981) and Huttenlocher and Lui (1979) described many differences between noun categories and verb categories, yet many accounts of research on concepts (including the present chapter) focus on noun categories. (For some recent examples of research on verb categories or event categories, especially in language acquisition, see Behrend, 1990; Kersten & Billman, 1992; Naigles, 1990; Nelson, 1986).

3. Questions of Structure

One anomalous observation, in our opinion, is that little attention has been directed to relational properties and their role in creating structure. For example, computational approaches to object recognition (e.g., Hummel & Biederman, 1992; Ullman, 1989) directly or indirectly incorporate structure. Given that EBL entails relational properties and structure, it might prove more feasible to integrate SBL and EBL if similarity models took a more structural view (e.g., Gentner, 1989; Goldstone and Medin, 1994a, 1994b). In addition to structural relations between features, we think that it will also be important to consider structural relations between category members. A concept is more than an unstructured collection of exemplars. For example, it is certainly true that priests, bishops, and nuns are members of the category *clergy,* but critical to understanding the *clergy* concept are the relations among these members. Some recent work has considered the relations between category members, suggesting that some categories have radial or chained structures (Lakoff, 1987; Malt, 1994).

4. Questions of Learning

Some of the most important research in categorization stems from the area of cognitive development. Observations on conceptual change lend important insight into categorization. Our review has notably failed to include a linguistic perspective on categorization, but surely language plays a critical role in determining what categories are learned and how they are learned (e.g., Landau & Gleitman, 1985; Pinker, 1992).

B. Opportunities

As we said earlier, we are guilty of optimism, and we think it's time for categorization research to become more ambitious. More ambitious in the sense that each of the perspectives we have reviewed (and others such as linguistics and computational vision) has methods, models, and insights that could effectively contribute in an integration of approaches. Each perspective can claim exciting developments, but each could be more powerful if it could borrow from its neighbors. We think it will happen.

Acknowledgments

Preparation of this chapter was supported by National Science Foundation grant 91-10245 to Douglas Medin. We are grateful to Lance Rips and Edward Wisniewski for comments on this chapter.

References

Ahn, W., Brewer, W. F., & Mooney, R. J. (1992). Schema acquisition from a single example. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 391–413.

Ahn, W., & Medin, D. L. (1992). A two-stage model of category construction. *Cognitive Science, 16,* 81–121.

Allen, S. W., & Brooks, L. R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General, 120,* 3–19.

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98,* 409–429.

Anglin, J. (1977). *Word, object, and conceptual development.* New York: Norton.

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 33–53.

Atran, S. (1987). Origins of the species and genus concepts. *Journal of the History of Biology, 20,* 195–279.

Atran, S. (1990). *Cognitive foundations of natural history: Towards an anthropology of science.* Cambridge, England: Cambridge University Press.

Atran, S. (1998). Folk biology and the anthropology of science, cognitive universals, and cultural particulars. *Behavioral and Brain Sciences, 21,* 547–609.

Bareiss, R. (1989). *Exemplar-based knowledge acquisition: A unified approach to concept representation, classification, and learning.* Boston: Academic Press.

Barkow, J. H., Cosmides, L., & Tooby, J. (1992). *The adapted mind: Evolutionary psychology and the generation of culture.* New York: Oxford University Press.

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition, 11,* 211–217.

Barsalou, L. W. (1985). Ideas, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 629–649.

Barsalou, L. W. (1990). On the indistinguishability of exemplar memory and abstraction in memory representation. In T. K. Srull & R. S. Wyer (Eds.), *Advances in social cognition* (pp. 61–88). Hillsdale, NJ: Erlbaum.

Barsalou, L. W. (1993). Structure, flexibility, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual systems. In A. C. Collins, S. E. Gathercole, & M. A. Conway (Eds.), *Theories of memory* (pp. 61–88). London: Erlbaum.

Behrend, D. A. (1990). The development of verb concepts: Children's use of verbs to label familiar novel events. *Child Development, 61,* 681–697.

Berlin, B. (1992). *Ethnobiological classification: Principles of categorization of plants and animals in traditional societies.* Princeton, NJ: Princeton University Press.

Berlin, B., Breedlove, D., & Raven, P. (1966). Folk taxonomies and biological classification. *Science, 154,* 273–275.

Berlin, B., Breedlove, D., & Raven, P. (1973). General principles of classification and nomenclature in folk biology. *American Anthropologist, 74,* 214–242.

Boster, J. S. (1986). Exchange of varieties and information between aguaruna manioc cultivators. *American Anthropologist, 88,* 429–436.

Boster, J. S. (1987). Agreement between biological classification systems is not dependent on cultural transmission. *American Anthropologist, 89,* 914–920.

Boster, J. S. (1988). Natural sources of internal category structure: Typicality, familiarity, and similarity of birds. *Memory & Cognition, 16,* 258–270.

Boster, J. S., Berlin, B., & O'Neill, J. (1986). The correspondence of Jivaroan to Scientific ornithology. *American Anthropologist, 88,* 569–586.

Boster, J. S., & D'Andrade, R. (1989). Natural and human sources of cross-cultural agreement in ornithological classification. *American Anthropologist, 91,* 132–142.

Boster, J. S., & Johnson, J. C. (1989). Form or function: A comparison of expert and novice judgments of similarities among fish. *American Anthropologist, 91,* 866–889.

Bourne, L. E., Jr. (1970). Knowing and using concepts. *Psychological Review, 77,* 546–556.

Boyer, P. (1990). *Tradition as truth and communication: A cognitive description of traditional discourse.* Cambridge, England: Cambridge University Press.

Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization.* pp. 169–211. New York: Wiley.

Brooks, L. R. (1987). Decentralized control of categorization: The role of prior processing episodes. In U. Neisser (Ed.), *Concepts and conceptual development: The ecological and intellectual factors in categorization* (pp. 141–174). Cambridge, England: Cambridge University Press.

Brooks, L. R., Norman, G. R., & Allen, S. W. (1991). Role of specific similarity in a medical diagnosis task. *Journal of Experimental Psychology: General, 120,* 278–287.

Busemeyer, J. R., Myung, I. J., & McDaniel, M. A. (1993). Theoretical implications for adaptive network learning models. *Psychological Science, 4,* 196–203.

Carey, S. A. (1985). *Conceptual change in childhood.* Cambridge, MA: Bradford Books.

Carey, S. A., & Gelman, R. (Eds.). (1991). *Epigenesis of mind: Studies in biology and cognition.* Hillsdale, NJ: Erlbaum.

Dejong, G. F. (1988). An introduction to explanation-based learning. In H. Shrobe (Ed.), *Exploring artificial intelligence* (pp. 45–81). San Mateo, CA: Morgan Kaufmann.

Dietterich, T. G. (1986). Learning at the knowledge level. *Machine Learning, 1,* 287–316.

Donnellan, K. S. (1977). Reference and definite descriptions. In S. P. Schwartz (Ed.), *Naming, necessity, and natural kinds* (pp. 42–65). Ithaca, NY: Cornell University Press.

Dupré, J. (1981). Natural kinds and biological taxa. *Philosophical Review, 90,* 66–90.

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inferences for psychological research. *Psychological Review, 70,* 193–242.

Ellman, T. (1989). Explanation-based learning: A survey of programs and perspectives. *Computing Surveys, 21,* 163–221.

Estes, W. K. (1993). Concepts, categories, and psychological science. *Psychological Science, 4,* 143–154.

Estes, W. K. (1994). *Classification and cognition* (pp. 161–163). Oxford, England: Oxford University Press.

Farah, M. J., & McClelland, J. L. (1991). A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General, 120,* 339–358.

Fisher, D. H., & Langley, P. (1990). The structure and formation of natural categories. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 241–284). San Diego, CA: Academic Press.

Fisher, D. H., Pazzani, M. J., & Langley, P. (Eds.). (1991). *Concept formation: Knowledge and experience in unsupervised learning.* San Mateo, CA: Morgan Kaufmann.

Fisher, D. H., & Yoo, J. P. (1993). Categorization, concept learning, and problem solving: A unifying view. In G. V. Nakamura, R. Taraban, & D. L. Medin (Eds.), *The psychology of learning and motivation: Categorization by humans and machines* (Vol. 29, pp. 219–255). San Diego, CA: Academic Press.

Flann, N. S., & Dietterich, T. G. (1989). A study of explanation-based methods for inductive learning. *Machine Learning, 4,* 187–226.

Flannagan, M. J., Fried, L. S., & Holyoak, J. K. (1986). Distributional expectations and the induction of category structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 241–256.

Fodor, J. A. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 234–257.

Garner, W. R. (1974). *The processing of information and structure.* Hillsdale, NJ: Erlbaum.

Gelman, R. (1990). First principles organize attention to and learning about relevant data: Number and the animate-inanimate distinction as examples. *Cognitive Science, 14,* 79–106.

Gelman, S. A., Coley, J. D., & Gottfried, G. M. (1994). Essentialist beliefs in children: The acquisition of concepts and theories. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind* (pp. 341–367). Cambridge, England: Cambridge University Press.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23,* 183–209.

Gelman, S. A., & Medin, D. L. (1993). What's so essential about essentialism? A different perspective on the interaction of perception, language, and conceptual knowledge. *Cognitive Development, 8,* 157–167.

Gelman, S. A., & Wellman, H. M. (1991). Insides and essence: Early understandings of the non-obvious. *Cognition, 38,* 213–244.

Gentner, D. (1981). Some interesting differences between verbs and nouns. *Cognition and Brain Theory, 4,* 161–178.

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 199–241). Cambridge, England: Cambridge University Press.

Gentner, D., & Ratterman, M. J. (1991). Language and the career of similarity. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspective on thought and language: Interrelations in development* (pp. 225–277). Cambridge, England: Cambridge University Press.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67.

Gluck, M. A. (1991). Stimulus generalization and representation in adaptive network models of category learning. *Psychological Science, 2,* 50–55.

Gluck, M. A., & Bower, G. H. (1988a). Evaluating an adaptive network model for human learning. *Journal of Memory and Language, 27,* 166–195.

Gluck, M. A., & Bower, G. H. (1988b). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General, 117,* 227–247.

Gluck, M. A., & Bower, G. H. (1990). Component and pattern information in adaptive networks. *Journal of Experimental Psychology: General, 119,* 105–109.

Goldstone, R. L., & Medin, D. L. (1994a). Similarity, interactive activation, and mapping: An overview. In K. Holyoak & J. Barnden (Eds.), *Advances in connectionist and neural computation theory: Connectionist approaches to analogy, metaphor, and case-based reasoning* (Vol. 2, pp. 321–362). Norwood, NJ: Ablex.

Goldstone, R. L., & Medin, D. L. (1994b). Time course of comparison. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 29–50.

Goodman, N. (1972). Seven strictures on similarity. In N. Goodman (Ed.), *Problems and projects* (pp. 437–447). New York: Bobbs-Merrill.

Grice, H. P. (1957). Meaning. *Philosophical Review, 66,* 377–388.

Hammond, K. J. (1989). *Case-based planning: Viewing planning as a memory task.* San Diego, CA: Academic Press.

Hampton, J. A. (1987). Inheritance of attributes in natural concept conjunctions. *Memory & Cognition, 15,* 55–71.

Hampton, J. A. (1988). Overextension of conjunctive concepts: Evidence for a unitary model of concept typicality and class inclusion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 12–32.

Hampton, J. A. (1993). Prototype models of concept representation. In I. V. Mechelen, J. Hampton, R. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive data analysis* (pp. 67–88). San Diego, CA: Academic Press.

Hastie, R., Schroeder, C., & Weber, R. (1990). Creating complex social conjunction categories from simple categories. *Bulletin of the Psychonomic Society, 28,* 242–247.

Hayes, B. K., & Taplin, J. E. (1992). Developmental changes in categorization processes: Knowledge and similarity-based models of categorization. *Journal of Experimental Child Psychology, 54,* 188–212.

Heit, E. (1992). Categorization using chains of examples. *Cognitive Psychology, 24,* 341–380.

Heit, E. (1993). Modeling the effects of expectations on recognition memory. *Psychological Science, 4,* 244–252.

Heit, E. (1994). Models of the effects of prior knowledge on category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1264–1282.

Heit, E., & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 411–422.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528–551.

Hintzman, D. L. (1991). Why are formal models useful in psychology? In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 39–56). Hillsdale, NJ: Erlbaum.

Hirschfeld, L. A. (1994). The child's representation of human groups. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 31, pp. 133–184). San Diego, CA: Academic Press.

Hirschfeld, L. A. (1995). Anthropology, psychology, and the meanings of social causality. In A. Premack, D. Premack, & D. Sperber (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 313–350). New York: Oxford University Press.

Hirschfeld, L. A., & Gelman, S. A. (1994). *Mapping the mind.* Cambridge, England: Cambridge University Press.

Horton, M. S., & Markman, E. M. (1980). Developmental differences in the acquisition of basic and superordinate categories. *Child Development, 51,* 708–719.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 99,* 480–517.

Huttenlocher, J., & Lui, F. (1979). The semantic organization of some simple nouns and verbs. *Journal of Verbal Learning and Verbal Behavior, 18,* 141–162.

James, W. (1890). *Principles of Psychology* (Vol. 1). New York: Holt.

Jones, C. M., & Heit, E. (1993). An evaluation of the total similarity principle: Effects of similarity on frequency judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 799–812.

Jones, S. S., & Smith, L. B. (1993). The place of perception in children's concepts. *Cognitive Development, 8,* 113–140.

Jones, S. S., & Smith, L. B., & Landau, B. (1991). Object properties and knowledge in early lexical learning. *Child Development, 62,* 499–516.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development.* Cambridge, MA: MIT Press.

Keil, F. C., & Batterman, N. (1984). A characteristic-to-defining shift in the development of word meaning. *Journal of Verbal Learning and Behavior, 23,* 221–236.

Kemler, D. G. (1982). Classification in young and retarded children: The primacy of overall similarity relations. *Child Development, 53,* 768–779.

Kemler, D. G. (1983). Exploring and reexploring issues of integrality, perceptual sensitivity, and dimensional salience. *Journal of Experimental Child Psychology, 36,* 365–379.

Kemler-Nelson, D. G. (1988). When category learning is holistic: A reply to Ward and Scott. *Memory & Cognition, 16,* 79–84.

Kersten, A. W., & Billman, D. O. (1992). *The role of correlational structure in learning event categories.* Proceedings of the 14th annual conference of the Cognitive Science Society, Bloomington, Indiana, pp. 432–437.

Kolodner, J. L., & Simpson, R. L. (1989). The MEDIATOR: Analysis of an early case-based problem solver. *Cognitive Science, 13,* 507–549.

Komatsu, L. K. (1992). Recent views of conceptual structure. *Psychological Bulletin, 112,* 500–526.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99,* 22–44.

Kunda, Z., Miller, D. T., & Claire, T. (1990). Combining social concepts: The role of causal reasoning. *Cognitive Science, 14,* 551–577.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago: University of Chicago Press.

Lamberts, K. (1994). Flexible tuning of similarity in exemplar-based categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1003–1021.

Landau, B., & Gleitman, L. R. (1985). *Language and experience: Evidence from the blind child.* Cambridge, MA: Harvard University Press.

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development, 3,* 299–321.

Leslie, A. M. (1984). Spatiotemporal continuity and the perception of causality in infants. *Perception, 13,* 287–305.

Losonksy, M. (1990). The nature of artifacts. *Philosophy, 65,* 81–89.

Malt, B. C. (1990). Features and beliefs in the mental representations of categories. *Journal of Memory and Language, 29,* 289–315.

Malt, B. C. (1994). Water is not H_2O. *Cognitive Psychology, 27,* 41–70.

Malt, B. C. (1995). Category coherence in cross-cultural perspectives. *Cognitive Psychology, 29,* 85–148.

Malt, B. C., & Smith, E. E. (1984). Correlated properties in natural categories. *Journal of Verbal Learning and Verbal Behavior, 23,* 250–269.

Mandler, J. M. (1992). How to build a baby: II. Conceptual primitives. *Psychological Review, 99,* 587–604.

Mandler, J. M. (1993). On concepts. *Cognitive Development, 8,* 141–148.

Mandler, J. M., Bauer, P. J., & McDonough, L. (1991). Separating the sheep from the goats: Differentiating global categories. *Cognitive Psychology, 23,* 263–298.

Mandler, J. M., & McDonough, L. (1993). Concept formation in infancy. *Cognitive Development, 8,* 291–318.

Margolis, E. (1994). A reassessment of the shift from the classical theory of concepts to prototype theory. *Cognition, 51,* 73–89.

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction.* Cambridge, MA: MIT Press.

Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science, 14,* 57–77.

Markman, E. M. (1992). Constraints on word learning: Speculations about their nature, origins, and domain specificity. In M. R. Gunnar & M. Maratsos (Eds.), *Modularity and constraints in language and cognition* (pp. 59–102). Hillsdale, NJ: Erlbaum.

Markman, E. M., & Wachtel, G. R. (1988). Children's use of mutual exclusivity to constrain the meaning of two words. *Cognitive Psychology, 20,* 121–157.

Marr, D. (1982). *Vision.* San Francisco: Freeman.

Martin, R. C., & Caramazza, A. (1980). Classification of well-defined and ill-defined categories: Evidence for common processing strategies. *Journal of Experimental Psychology: General, 109,* 320–353.

Matheus, C. J., Rendell, L. R., Medin, D. L., & Goldstone, R. C. (1989). *Purpose and conceptual functions: A framework for concept representation and learning in humans and machines.* Proceedings of the 7th annual conference of the Society for the Study of Artificial Intelligence and Simulation of Behavior. T. Cohn (Ed.). Sussex, England: Pitmen Publishing.

Mayr, E. M. (1982). *The growth of biological thought: Diversity, evolution, and inheritance.* Cambridge, MA: Harvard University Press.

McClelland, J. L., & Rumelhart, D. E. (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition: Psychological and biological models* (Vol. 2). Cambridge, MA: MIT Press.

Medin, D. L. (1986). Comment on "Memory storage and retrieval processes in category learning." *Journal of Experimental Psychology: General, 115,* 373–381.

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist, 44,* 1469–1481.

Medin, D. L., Altom, M. W., Edelson, S. M., & Freko, D. (1982). Correlated symptoms and simulated medical classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 37–50.

Medin, D. L., & Edelson, S. E. (1988). Problem structure and the use of base-rate information from experience. *Journal of Experimental Psychology: General, 117,* 68–85.

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning.* Cambridge, England: Cambridge University Press.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207–238.

Medin, D. L., & Schwanenflugel, P. J. (1981). Linear separability in classification learning. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 355–368.

Medin, D. L., & Shoben, E. J. (1988). Context and structure in conceptual combination. *Cognitive Psychology, 20,* 158–190.

Medin, D. L., & Smith, E. E. (1984). Concepts and concept formation. *Annual Review of Psychology, 35,* 113–138.

Medin, D. L., & Thau, D. M. (1992). Theories, constraints, and cognition. In H. L. Pick, P. Van Den Broek, & D. C. Knill (Eds.), *Cognition: Conceptual and methodological issues* (pp. 165–188). Washington, DC: American Psychological Association.

Medin, D. L., Wattenmaker, W. D., & Michalski, R. S. (1987). Constraints and preferences in inductive learning: An experimental study of human and machine performance. *Cognitive Science, 11,* 299–339.

Mervis, C. B. (1987). Child-basic object categories and early lexical development. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 201–233). New York: Cambridge University Press.

Mervis, C. B., Johnson, K. E., & Scott, P. (1993). Perceptual knowledge, conceptual knowledge, and expertise: Comment on Jones and Smith. *Cognitive Development, 8,* 149–155.

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology, 32,* 89–115.

Michalski, R. S. (1983a). A theory and methodology of inductive learning. *Artificial Intelligence, 20,* 111–161.

Michalski, R. S. (1983b). A theory and methodology of inductive learning. *Artificial Intelligence, 2,* 111–163.

Michalski, R. S. (1993). Inferential learning theory as a conceptual basis for multistrategy learning. *Machine Learning, 11,* 111–151.

Michalski, R. S., Mozetic, I., Hong, J., & Lavrac, N. (1986). The multipurpose incremental learning system AQ15 and its testing application to three medical domains. In *Proceedings of the Fifth Annual Conference on Artificial Intelligence* (pp. 1041–1045). Philadelphia: Morgan Kaufmann.

Minsky, M. L., & Papert, S. A. (1988). *Perceptrons.* Cambridge, MA: MIT Press.

Mitchell, T. M., Keller, R., & Kedar-Cabelli, S. (1986). Explanation-based generalization: A unifying view. *Machine Learning, 10,* 79–110.

Mooney, R. J. (1993). Integrating theory and data in category learning. In G. V. Nakamura, R. Taraban, & D. L. Medin (Eds.), *The psychology of learning and motivation: Categorization by humans and machines* (Vol. 29, pp. 189–219). San Diego, CA: Academic Press.

Murphy, G. L. (1988). Comprehending complex concepts. *Cognitive Science, 12,* 529–562.

Murphy, G. L. (1993). Theories and concept formation. In I. V. Mechelen, J. Hampton, R. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive data analysis* (pp. 173–200). London: Academic Press.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289–316.

Murphy, G. L., & Wisniewski, E. J. (1989). Categorizing objects in isolation and in scenes: What a superordinate is good for. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 572–586.

Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language, 17,* 357–374.

Nakamura, G. V., Taraban, R., & Medin, D. L. (Eds.). (1993). *The psychology of learning and motivation: Categorization by humans and machines* (Vol. 29). San Diego, CA: Academic Press.

Nelson, K. (1986). *Event knowledge: Structure and function in development.* Hillsdale, NJ: Erlbaum.

Nosofsky, R. M. (1986). Attention, similarity and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115,* 39–57.

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 87–108.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 700–708.

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition in memory. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 3–27.

Nosofsky, R. M. (1992). Exemplar-based approach to relating categorization, identification, and recognition. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 363–393). Hillsdale, NJ: Erlbaum.

Nosofsky, R. M., & Kruschke, J. K. (1992). Investigations of an exemplar-based connectionist model of category learning. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 28, pp. 207–250). New York: Academic Press.

Nosofsky, R. M., Krushke, J. K., & McKinley, S. C. (1992). Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 211–233.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101,* 53–79.

Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97,* 185–200.

Osherson, D. N., Stern, J., Wilkie, O., Stob, M., & Smith, E. E. (1991). Default probability. *Cognitive Science, 15,* 251–269.

Pazzani, M. J. (1991). Influence of prior knowledge on concept acquisition: Experimental and computational results. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 416–432.

Pinker, S. (1992). Language and species. *Language, 68,* 375–383.

Porter, B. W., Bareiss, R., & Holte, R. C. (1990). Concept learning and heuristic classification in weak-theory domains. *Artificial Intelligence, 45,* 229–263.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77,* 353–363.

Premack, A. (Ed.). (1994). *Causal understanding in cognition and culture.* New York: Oxford University Press.

Quine, W. V. O. (1960). *Word and object.* Cambridge, MA: MIT Press.

Rajamoney, S. A. (1990). A computational approach to theory revision. In J. Shrager & P. Langley (Eds.), *Computational models of scientific discovery and theory formation* (pp. 225–254). San Mateo, CA: Morgan Kaufmann.

Reagher, G., & Brooks, L. R. (1993). Perceptual manifestations of an analytic structure: The priority of holistic individuation. *Journal of Experimental Psychology: General, 122,* 92–114.

Rendell, L. A., & Cho, H. (1990). Empirical learning as a function of concept character. *Machine Learning, 5,* 267–298.

Rey, G. (1983). Concepts and stereotypes. *Cognition, 15,* 237–262.

Richman, H. (1991). Discrimination net models of concept formation. In D. Fisher, M. Pazzani, & P. Langley (Eds.), *Concept formation: Knowledge and experience in unsupervised learning* (pp. 103–125). San Mateo, CA: Morgan Kaufmann.

Riesbeck, C. K., & Schank, R. C. (1989). *Inside case-based reasoning.* Hillsdale, NJ: Erlbaum.

Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior, 14,* 665–681.

Rips, L. J. (1989). Similarity, typicality and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21–59). Cambridge, England: Cambridge University Press.

Rips, L. (1995). The current status of research on conceptual combination. *Mind and Language, 10,* 72–104.

Romney, A. K., Weller, S. C., & Batchelder, W. H. (1986). Culture as consensus: A theory of culture and informant accuracy. *American Anthropologist, 88,* 318–338.

Rosch, E. (1973). Natural categories. *Cognitive Psychology, 4,* 328–350.

Rosch, E. (1978). Principles of categorization. In E. Rosch, & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Erlbaum.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology,* 7, 573–605.

Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 573–605.

Ross, B. H., & Kennedy, P. T. (1990). Generalizing from the use of earlier examples in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 42–55.

Ross, B. H., Perkins, S. J., & Tenpenny, P. L. (1990). Reminding-based category learning. *Cognitive Psychology, 22,* 460–492.

Rothbart, M., & Taylor, M. (1992). Category labels and social reality: Do we view social categories as natural kinds? In G. Semin & K. Fielder (Eds.), *Language, interaction and social cognition* (pp. 11–36). Beverly Hills, CA: Sage.

Rumelhart, D. E., & McClelland, J. R. (Eds.). (1986). Parallel distributed processing. *Explorations in the microstructure of cognition: Foundations* (Vol. 1). Cambridge, MA: MIT Press.

Rumelhart, D. E., & Zipser, D. (1985). Feature discovery by competitive learning. *Cognitive Science, 19,* 75–112.

Sartori, G., & Job, R. (1988). The oyster with four legs: A neuropsychological study on the interaction of visual and semantic information. *Cognitive Neuropsychology, 15,* 105–132.

Schank, R. C., Collins, G. C., & Hunter, L. E. (1986). Transcending inductive category formation in learning. *Behavioral and Brain Sciences, 9,* 639–651.

Schwanenflugel, P. J., & Rey, M. (1986). Interlingual semantic facilitation: Evidence for a common representational system in the bilingual lexicon. *Journal of Memory and Language, 25,* 605–618.

Schwartz, S. P. (1977). *Naming, necessity, and natural kinds.* Ithaca, NY: Cornell University Press.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237,* 1317–1323.

Shepp, B. E., & Swartz, K. B. (1976). Selective attention and the processing of integral and non-integral dimensions: A developmental study. *Journal of Experimental Child Psychology, 22,* 73–85.

Shipley, E. F. (1993). Categories, hierarchies, and induction. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 30, pp. 265–301). San Diego, CA: Academic Press.

Smith, E. E., & Medin, D. L. (1981). Categories and concepts. Cambridge, MA: Harvard University Press.

Smith, E. E., Medin, D. L., & Rips, L. J. (1984). A psychological approach to concepts: Comments on Rey's "Concepts and stereotypes." *Cognition, 17*, 265–274.

Smith, E. E., & Osherson, D. N. (1984). Conceptual combination with prototype concepts. *Cognitive Science, 8*, 337–361.

Smith, E. E., Osherson, D. N., Rips, L. J., & Keane, M. (1988). Combining prototypes: A selective modification model. *Cognitive Science, 12*, 485–527.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review, 81*, 214–241.

Smith, E. E., & Sloman, S. (1994). Similarity-versus rule-based categorization. *Memory & Cognition, 22*, 377–386.

Smith, E. R. (1990). Content and process specificity in the effects of prior experiences. In T. K. Srull & R. S. Wyer, Jr. (Eds.), *Advances in social cognition* (pp. 1–59). Hillsdale, NJ: Erlbaum.

Smith, E. R., & Zarate, M. A., Exemplar-based models of social judgment. *Psychological Review, 99*, 3–21.

Smith, L. B. (1989). A model of perceptual classification in children and adults. *Psychological Review, 96*, 125–144.

Smith, L. B., & Kemler, D. G. (1977). Developmental trends in free classification: Evidence for a new conceptualization of perceptual development. *Journal of Experimental Child Psychology, 24*, 279–298.

Smith, L. B., & Kemler, D. G. (1978). Levels of experienced dimensionality in children and adults. *Cognitive Psychology, 10*, 502–532.

Spalding, T. L., & Ross, B. H. (1994). Comparison-based learning: Effects of comparing instances during category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1251–1263.

Springer, K., & Keil, F. C. (1991). Early differentiation of causal mechanisms appropriate to biological and nonbiological kinds. *Child Development, 62*, 767–785.

Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology, 23*, 457–482.

Thau, D. M. (1992). *Primacy effects and selective attention incremental clustering.* Proceedings of the 14th annual conference of the Cognitive Science Society, Bloomington, Indiana, pp. 219–223.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327–352.

Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition, 32*, 193–254.

Walker, S. J. (1992). Supernatural beliefs, natural kinds, and conceptual structure. *Memory & Cognition, 20*, 655–662.

Ward, T. B., Vela, E., & Hass, S. D. (1990). Children and adults learn family-resemblance categories analytically. *Child Development, 61*, 393–406.

Warrington, E. K., & McCarthy, R. (1983). Category specific access dysphasia. *Brain, 106*, 859–878.

Warrington, E. K., & McCarthy, R. (1987). Categories of knowledge: Further fractionation and an attempted integration. *Brain, 110*, 1237–1296.

Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain, 107*, 829–854.

Wattenmaker, W. D. (1995). Knowledge structures and linear separability: Integrating information in object and social categorization. *Cognitive Psychology, 28*, 273–328.

Wattenmaker, W. D., Dewey, G. I., Murphy, T. D., & Medin, D. L. (1986). Linear separability and concept learning: Context, relational properties, and concept naturalness. *Cognitive Psychology, 18*, 158–194.

Weber, E. U., Bockenholt, U., Hilton, D. J., & Wallace, B. (1993). Determinants of diagnostic hypothesis generation: Effects of information, base rates, and experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1131–1165.

Weller, S. C. (1987). Shared knowledge, intracultural variation, and knowledge aggregation. *American Behavioral Scientist, 31*, 178–193.

Whittlesea, B. W. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1235–1254.

Wisniewski, E. J., & Medin, D. L. (1991). Harpoons and long sticks: The interaction of theory and similarity in rule induction. In D. H. Fisher, M. J. Pazzani, & P. Langley (Eds.), *Concept formation: Knowledge and experience in unsupervised learning* (pp. 237–278). San Mateo, CA: Morgan Kaufmann.

Wisniewski, E. J., & Medin, D. L. (1994a). The fiction and nonfiction of features. In R. S. Michalski & G. Tecuci (Eds.), *Machine learning* (Vol. 4, pp. 63–84). San Mateo, CA: Morgan Kaufmann.

Wisniewski, E. J., & Medin, D. L. (1994b). On the interaction of theory and data in concept learning. *Cognitive Science, 18,* 221–281.

Yoo, J., & Fisher, D. H. (1991). Concept formation over problem-solving experience. In D. H. Fisher, M. J. Pazzani, & P. Langley (Eds.), *Concept formation: Knowledge and experience in unsupervised learning* (pp. 279–303). San Mateo, CA: Morgan Kaufmann.

Yuill, N. (1992). Children's conception of personality traits. *Human Development, 35,* 265–279.